8-1-1977

# A survey of economic forecasting techniques

Eugene Michael Boyd
*University of Nebraska at Omaha*

# A SURVEY OF ECONOMIC FORECASTING TECHNIQUES

A Thesis

Presented to the

Department of Economics

and the

Faculty of the Graduate College

University of Nebraska

In Partial Fulfillment

of the Requirements for the Degree

Master of Arts

University of Nebraska at Omaha

by

Eugene Michael Boyd

August, 1977

THESIS ACCEPTANCE

Accepted for the faculty of the Graduate College, University of Nebraska, in partial fulfillment of the requirements for the degree Master of Arts, University of Nebraska at Omaha.

THESIS COMMITTEE

| Name | Department |
|---|---|
| James Jorg | Economics |
| Keith K. Turner | Economics |
| Donald A. Nielsen | Real Estate & Land Use Econo. |

David L. Schulz — Economic.

Chairman

July 14, 1977

Date

TABLE OF CONTENTS

# CHAPTER I

## INTRODUCTION

### A. Economic Forecasting

Decision making involves analysis of relevant factors in order to determine a course of action. Since the action to be taken carries into the future, a decision must take into consideration the future status of and effects of any changes in these relevant factors. If these factors are likely to change in the future, the effect of any change must be considered now, before the decision is made. The problem is that it is difficult to foresee these changes. Because the future cannot be known perfectly, and because the correctness of a decision depends on the future status of relevant factors, the decision-makers must often rely on a forecast. As Theil puts it quite simply, "A forecast or a prediction...is defined as a statement concerning unknown, particularly future, events."[1]

In planning a course of action, the decision-maker seeks to eliminate as much uncertainty as possible. Forecasts based on methods with proven reliability increase the confidence with which decisions may be made.

---

[1] Henri Theil, Applied Economic Forecasting, (Amsterdam: North-Holland, 1966), p.1.

Forecasting may take many forms and be used in many different activities, such as agriculture, industry, government, etc. Economic forecasting is concerned with eliminating some of the uncertainty surrounding the future course of such economic variables as sales, income, or employment. A forecast of future sales allows a firm to better plan its production. Forecasts of such macroeconomic variables as income and employment are needed in order for the federal government to plan policy and budgets.

A study of economic variables and their underlying relationships and interdependence via statistical procedures allows the forecasting of these variables with some degree of confidence. Depending upon the forecast method used, this confidence may be based on formal statistical measures of probability, significance, and confidence intervals, or on empirical evidence of accuracy from past observations.

A forecast may be in the form of a point prediction or of an interval estimation. A point prediction, as might be given by a naive trend projection method, states a predicted value for the variable for some future time period. Nothing can be said about the formal statistical characteristics of such a forecast. The confidence in the forecast is based on past accuracy of the method. More sophisticated methods, such as regression analysis, allow statistical analysis of the probability of the variable falling within a given range or interval. For example, in addition to the point estimation given by such a method, we can also compute the standard deviation, and know that there is a 95% probability that the true value will be that point plus or minus 1.96 standard deviations.

Forecasting is therefore a useful tool in dealing with the uncertainty of decision-making.

## B. Forecasting vs. Estimation

In the theoretical study of economic variables, the economist wishes to discover all relationships relevant to the subject under consideration, and the strength of those relationships. The process of estimation involves the development of a function which attempts to describe the nature and strength of the underlying relationships between a given variable and other determining factors. These relationships are generally formulated through the use of economic theory and then tested through the use of statistical or econometric methods. Dutta states that

> ...the essential contribution of econometrics is constructing models that provide tests of the agreement between theoretical formulations and empirical observations.[1]

For example, an attempt to estimate a demand function for a particular commodity might involve an examination of the price of the commodity, the prices of substitute and complementary goods, the general level of income, the effects of advertising, and so on. From demand theory and

---

[1]M. Dutta, Econometric Methods (Cincinnati: South-Western, 1975), p. 2.

from empirical observations of the variables, a demand function would then be formulated. As stated by Elliott,

> ...a demand function is an attempt to quantify the relationship between influential variables such as price and the associated quantity buyers will want....(The function) formalizes the factors that together explain the amount buyers will want to buy.[1]

This demand function or model is assumed to be an approximation of the underlying relationships between the variables. Dutta states this as follows.

> A model offers a formal description or the relationships between endogenous and exogenous variables. A structure is defined by a "true" description of the same set of relationships, in terms of the "true" values of the parameters...."True" structural parameters are seldom known....The numerical values of the variables observed are in fact believed to be generated by the "true" structure.[2]

From the practical standpoint of the decision-maker, such an all-inclusive study may not be necessary. Indeed, it may be undesirable due to excessive costs and/or time needed to complete such a study.

What is desired then is not an estimation, but a forecast. A forecast may be based on some of these underlying relations, on one of then, or may not attempt to use them at all. A forecast is generally desired for its accuracy and usefulness, not for the knowledge it may provide about the structure of underlying economic relationships. If a simple technique with no explicit reference

---

[1]J. Walter Elliott, Economic Analysis for Management Decisions (Homewood, Ill.: Irwin, 1973) p. 100.

[2]Dutta, op. cit., pp. 12-13.

to structure or to any relationship between variables seems to provide the best forecast of the future, it will be the method which is used. According to Joel Dean,

> Pragmatic analysis usually has a much more limited objective than an entire demand function, because for a single purpose, the best information comes when attention is concentrated on the few demand determinants most important for that purpose.
>
> A theoretical demand function strives for comprehensiveness; it usually seeks to include all the forces that may influence sales. For empirical analysis, the aim should be to select from the many demand determinants only those that are variable, important, measurable, independent, and (for some problems) controllable...[1]

The added accuracy and confidence derived from the inclusion and study of additional variables must be weighed against the increased time and expense thereby incurred.

Depending on the nature of the problem being examined, it may be possible to make confident forecasts on the basis of one series without including any determining or exogenous variables. Various methods of simple forecasts, such as simple trend projection, allow forecasts to be made solely on the basis of the series being studied.

> Economic time series do, for the most part, show a persistent tendency to move in the same direction for a period of time because of their inherent cumulative characteristics.[2]

Although other, determining, variables are not directly considered in using such forecasting techniques, the effects of these variables are present insofar as they have helped

---

[1]Joel Dean, Managerial Economics (Englewood Cliffs: Prentice Hall, 1951), p. 165.

[2]Milton H. Spencer and Louis Siegelman, Managerial Economics (Homewood: Irwin, 1959), p. 29.

to determine the past values of the series. The presence of underlying relationships may thus be utilized without quantification of such relationships, and even without specific knowledge of them.

## C. Methodology

In subsequent chapters, various forecasting techniques will be examined. Chapter II will look at the so-called "naive" methods, which forecast without reference to other economic variables. These methods are generally of the trend projection type, involving an analysis of patterns within a series and using any such patterns to predict future values for the series.

> Naive methods of forecasting are typically unsophisticated and unscientific projections based on guesses or on mechanical extrapolations of historic data....Typically, they are distinguished from other forecasting methods...in that they are essentially mechanical and are not closely integrated with relevant economic theory and statistical data. [1]

Naive forecasting techniques, in not taking other economic variables into consideration, ignore underlying structure except to the extent that such structure has manifested itself in determining past values of the series. Also, being simple extrapolation methods, they give no statistical information, such as standard errors or confidence intervals, about the forecasts. Nevertheless, they can give useful forecasts for certain types of series, depending on the nature of the series.

---

[1] Spencer and Siegelman, op. cit., p.26.

Chapter III will deal with economic indicators. These methods involve the examination of the behavior of one or more economic series in order to predict the behavior of some other series. Indicator techniques are generally non-quantitative in nature, and are more concerned with direction than with magnitude of change.

Chapter IV will be concerned with regression techniques, and Chapter V with their application to forecasting. This concerns the use of econometric-type methods for forecasting. From the more limited viewpoint of forecasting...

> ...econometrics...is a method of explaining past economic activity, and of predicting future economic activity, by deriving mathematical equations that will express the most probable interrelationship between a set of variables....By combining the relevant variables...into what seems to be the best mathematical arrangement, econometricians proceed to predict the future course of one or more of these variables on the basis of the established relationships. The "best mathematical arrangement" is thus a model...that seems to best describe the past set of relationships according to economic theory and statistical analysis.[1]

Treatment of each forecasting technique will include the theory and formulation of the particular method, together with any statistical problems and tests associated with it. The technique will then be illustrated by application to the problem of forecasting the mortgage credit series. The mortgage series has been selected because it is an important economic variable, affecting much

---

[1]Spencer and Siegelman, op. cit., p.40.

of government and business. Mortgages, and the concomitant demand for housing, affect government planning ranging from the provision of local services to federal policy in many areas. In the private sector, finacial intermediaries, building contractors, producers of consumer durables, and many others are either directly or indirectly consermed with activity in this area.

The series used will be outstanding mortgage debt held by saving and loan associations, as reported in the Federal Reserve Bulletin. Quarterly "forecasts" will be made for the eight periods from 1975:1 to 1976:4, based on previous data. The number of previous periods included in computing the forecast will vary according to the nature of the method being used. These "forecasts" will then be compared with the observed values for the eight periods, and accuracy measurements will be made through the use of two statistical methods.

The first accuracy measurement used is the root-mean-square forecast error. This is computed as

$$\text{RMS} = \sqrt{\frac{1}{n} \sum (P_i - A_i)^2}$$

where $P_i$ is the forecast value and $A_i$ the actual value for each period. The RMS measurement deals in absolute terms, giving an average forecast error in the units of the series being forecast. This allows various methods to be compared for accuracy and gives a concrete idea of the amount of error involved in the forecast.

The second accuracy measurement used is Theil's Inequality Coefficient, $U^2$ (or sometimes designated $I^2$). This is computed by the formula

$$U^2 = \frac{\sum (P_i - A_i)^2}{\sum A_i^{\,2}}$$

where $P_i$ is the predicted percent change in the variable and $A_i$ is the actual percent change. The lower limit of $U^2$ is zero, occurring when all predicted values are perfect forecasts, so that $P_i = A_i$ for all periods forecast. While there is no upper limit on the value which $U^2$ may take, the pragmatic limitation is the value of $U^2 = 1$. If a simple forecast of no change is made for all periods, so that the predicted percent change is zero ($P_i = 0$ for all forecast periods), then $U^2 = 1$. Therefore, any forecasting method resulting in a value of $U^2$ greater than one may be considered to be of less use than no forecast at all. This measurement then, while allowing the comparison of various methods, also gives a standard by which the results may be evaluated. While one forecasting method may give better results than another, the $U^2$ may show that neither is really acceptable, and that other methods should be tried.

Following the accuracy measurements, the applicability of each technique will be discussed, bringing into consideration the nature of the technique and the nature of the series.

The final chapter will summarize the results and give a general evaluation of the various methods.

# CHAPTER II

## NAIVE METHODS

### A. Continuity Method

The continuity method is a simple forecast that there will be no change in the value of the variable in the forecast period. Therefore, the predicted value for the forecast period is equal to the variable value for the current period. A forecast of this type is based on the assumption that the series is being generated by a stochastic process, that each value of Y is drawn randomly from a probability distribution. Since the probability distribution of a time series is not generally known, futher assumptions about the series must be made in order to forecast the series.

A continuity forecast assumes that the series under consideration follows a random walk process. In such a process, each value of the variable is drawn independently from a probability distribution with zero mean. In a random walk series, the expected value for a future period is equal to the current value plus some error term. Since the expected value of this error term is its mean value, which is zero, the forecast of the variable for any future period is the current value of the variable.

Continuity forecasts for the mortgage series are as follows. In the table below, $Y_p$ is the predicted value, $Y_a$ the actual value, $P_i$ the predicted percent change, and $A_i$ the actual percent change. (In subsequent tables, the values for $Y_a$ and $A_i$ will not be repeated, as they remain constant.) RMS and $U^2$ are calculated according to the formulas given in Chapter I, Section C.

TABLE 1

CONTINUITY METHOD

| | $Y_p$ | $Y_a$ | $Y_p-Y_a$ | $P_i$ | $A_i$ | $P_i-A_i$ |
|---|---|---|---|---|---|---|
| 1975:1 | 249.3 | 252.4 | - 3.1 | 0 | 1.24 | -1.24 |
| 2 | 252.4 | 261.3 | - 8.9 | 0 | 3.53 | -3.53 |
| 3 | 261.3 | 270.6 | - 9.3 | 0 | 3.56 | -3.56 |
| 4 | 270.6 | 278.7 | - 8.1 | 0 | 2.99 | -2.99 |
| 1976:1 | 278.7 | 286.6 | - 7.9 | 0 | 2.83 | -2.83 |
| 2 | 286.6 | 299.6 | -13.0 | 0 | 4.54 | -4.54 |
| 3 | 299.6 | 312.1 | -12.5 | 0 | 4.17 | -4.17 |
| 4 | 312.1 | 323.2 | -11.1 | 0 | 3.56 | -3.56 |

RMS = 9.69 $U^2 = 1$

As can be seen from these forecasts, the continuity or random walk method does not give good results when applied to the mortgage series. A cursory inspection of the series indicates the presence of a persistent upward trend. In the presence of such a trend, the continuity or random walk method is not applicable. In a series with a persistent upward trend, the difference between the value of one period and that of the next generally can be expected to be positive. Such knowledge about future occurances violates the random walk assumption of a random probability distribution with zero mean. If a series moves in one direction over time, its mean changes over time. The continuity method is better suited for forecasting a stable series whose mean is constant over time, with truly random changes both in magnitude and in direction. Essentially, this is a method of reaching some useful value for a variable when no better method is applicable and no futher information is available. Also, it provides a standard against which other forecasting techniques can be compared. Since the continuity method is actually a case of making no forecast, any method which does not give better results than this method can be considered to be worse than no forecast at all.

## B. Moving Average Method

The moving average method of forecasting is a deterministic model; that is, no reference is made to the nature of the underlying randomness in the series. The forecast value is the mean of a given number of observations of the variable, using only the most recent observations rather than the entire series. The forecast is given by

$$Y_{t+1} = \frac{1}{n} (Y_t + Y_{t-1} + \ldots + Y_{t-n-1})$$

The choice of the value of 'n' depends upon the data being used and the importance of earlier observations in determining the variable. For a simple forecast, enough observations should be included to remove any seasonality in the data. Seasonality may be tested and allowed for through the use of other techniques covered later. A first approximation for a moving average method might be n=12 for monthly data, and n=4 for quarterly data. Table 2 shows a simple moving average method using n=4.

TABLE 2

MOVING AVERAGE

| | $Y_p$ | $Y_p - Y_a$ | $P_i$ | $P_i - A_i$ |
|---|---|---|---|---|
| 1975:1 | 244.3 | - 8.1 | -2.00 | -3.24 |
| 2 | 248.3 | -13.0 | -1.62 | -5.15 |
| 3 | 252.7 | -17.9 | -3.29 | -6.85 |
| 4 | 258.4 | -20.3 | -4.51 | -7.50 |
| 1976:1 | 265.8 | -20.8 | -4.63 | -7.46 |
| 2 | 274.3 | -25.3 | -4.29 | -8.83 |
| 3 | 283.9 | -28.2 | -5.24 | -9.41 |
| 4 | 294.2 | -29.0 | -5.73 | -9.29 |
| | RMS = 21.44 | | $U^2$ = 4.758 | |

The use of a moving average method constantly updates the observations being considered, adding new data as it becomes available, and dropping earlier observations. If more recent observations are assumed to be more important in determining the current value of a variable than are the earlier values retained in the calculation, these more recent values can be weighted more heavily in computing the average. This can be done via an exponentially weighted moving average (EWMA). This is computed as

$$Y_{t+1} = aY_t + a(1-a)Y_{t-1} + a(1-a)^2Y_{t-2} + \ldots$$

The value of 'a' is between 0 and 1. The EWMA method gives more importance to the more recent observations of the variable, this importance increasing as the value of 'a' approaches 1. The value of 'a' is chosen via an iterative method, using that value which gives the most accurate results. For example, the EWMA may be computed for a=0.5 and for a=0.9, with results compared with actual observations. If results when a=0.9 are closer to actual observations, the average would then be computed for a=0.7, and for a=0.8. The iterative process thus narrows down the value of 'a' until the most accurate value for the series under consideration is found, based on past observations. This value is then used to make forecasts for future periods. In Table 3, the EWMA is computed with a=0.9.

TABLE 3

EXPONENTIALLY WEIGHTED MOVING AVERAGE

| | $Y_p$ | $Y_p-Y_a$ | $P_i$ | $P_i-A_i$ |
|---|---|---|---|---|
| 1975:1 | 249.1 | - 3.3 | -0.08 | -1.32 |
| 2 | 252.0 | - 9.3 | -0.16 | -3.69 |
| 3 | 260.4 | -10.2 | -0.34 | -3.90 |
| 4 | 269.6 | - 9.1 | -0.37 | -3.36 |
| 1976:1 | 277.8 | - 8.8 | -0.32 | -3.15 |
| 2 | 285.7 | -13.9 | -0.31 | -4.85 |
| 3 | 298.2 | -13.9 | -0.47 | -4.64 |
| 4 | 310.7 | -12.5 | -0.45 | -4.01 |

RMS = 10.63 $U^2$ = 1.197

While the EWMA method gives more accurate forecasts than the simple moving average method, neither is as accurate as the continuity method. A moving average method is handicapped in forecasting any series containing a persistent trend because this method gives greater weight to earlier observations. As can be seen in the EWMA forecasts, accuracy in forecasting this series improves as more weight is given to more recent observations, due to the strong upward trend. However, no matter how much weight is given to recent values, a forecast of a series with a strong upward trend can only approach the most recent value; it cannot be higher than that value.

Like the continuity method, a moving average technique, either simple or weighted, is better for forecasting a series exhibiting stationary characteristics, one whose mean does not vary over time. A series with short-run fluctuations about a fairly stable mean over the long-run could be more accurately forecast with these methods than a series showing a definite long-run trend.

## C. Trend Projection

A simple trend projection technique looks at very recent changes in a series and makes the forecast that this recent change will occur again. The change which is projected may be the absolute amount of change which recently occurred , or it may be the recent rate of change. The projected change may be that of the most recent period, or it may be the average of a number of past periods.

A trend projection of the absolute change of one period is given by the formula

$$Y_{t+1} = Y_t + (Y_t - Y_{t-1})$$

and a one period projection of rate of change is given by

$$Y_{t+1} = Y_t(Y_t \;/\; Y_{t-1})$$

The absolute change method simply forecasts that the series will increase (decrease) by the same amount that it increased (decreased) in the previous period. The rate of change method projects that the series will change by the same percent and in the same direction that it did in the previous period.

TABLE 4

ABSOLUTE CHANGE--ONE PERIOD

| | $Y_p$ | $Y_p - Y_a$ | $P_i$ | $P_i - A_i$ |
|---|---|---|---|---|
| 1975:1 | 251.0 | -1.4 | 0.68 | -0.59 |
| 2 | 255.5 | -5.8 | 1.23 | -2.30 |
| 3 | 270.2 | -0.4 | 3.41 | -0.15 |
| 4 | 279.9 | 1.2 | 3.44 | 0.45 |
| 1976:1 | 286.8 | 0.2 | 2.91 | 0.08 |
| 2 | 294.5 | -5.1 | 2.76 | -1.78 |
| 3 | 312.6 | 0.5 | 4.34 | 0.17 |
| 4 | 324.6 | 1.4 | 4.00 | 0.44 |

RMS = 2.86 $U^2$ = 0.098

Forecasts based on the rate of change method for one period are given below in Table 5. Either of these simple trend projection methods may be expanded to take into consideration the changes of a number of past periods. By considering more than just the most recent period, any seasonality or irregular influence in the series can be averaged out. A projection of the absolute change of a number of recent periods (n-periods) is given by the formula:

$$Y_{t+1} = Y_t + \frac{\sum_{i=0}^{n} (Y_{t-i} - Y_{t-i-1})}{(n+1)}$$

Forecasts using this technique are given in Table 6.

The trend projection of the rate of change for n-periods is computed by

$$Y_{t+1} = Y_t \left[ \sum_{i=0}^{n} \frac{Y_{t-i}}{Y_{t-i+1}} \right] \left[ \frac{1}{(n+1)} \right]$$

Forecasts based on this method are given in Table 7.

Accuracy measurements for these methods show great improvement over previously discussed techniques. The root-mean-square forecast error is greatly reduced, and the $U^2$ value is much lower than the $U^2 = 1$ of the continuity method, the pragmatic upper limit of this test.

TABLE 5

RATE OF CHANGE--ONE PERIOD

| | $Y_p$ | $Y_p - Y_a$ | $P_i$ | $P_i - A_i$ |
|---|---|---|---|---|
| 1975:1 | 251.0 | -1.4 | 0.68 | -0.56 |
| 2 | 255.5 | -5.8 | 1.23 | -2.30 |
| 3 | 270.5 | -0.1 | 3.52 | -0.04 |
| 4 | 280.2 | 1.5 | 3.66 | 0.67 |
| 1976:1 | 287.0 | 0.4 | 2.98 | 0.15 |
| 2 | 294.7 | -4.9 | 2.82 | -1.72 |
| 3 | 313.2 | 1.1 | 4.54 | 0.37 |
| 4 | 325.1 | 1.9 | 4.16 | 0.60 |

RMS = 2.89 $U^2$ = 0.101

TABLE 6

ABSOLUTE CHANGE--N-PERIODS (N=3)

| | $Y_p$ | $Y_p-Y_a$ | $P_i$ | $P_i-A_i$ |
|---|---|---|---|---|
| 1975:1 | 253.8 | 1.4 | 1.80 | 0.56 |
| 2 | 256.7 | -4.6 | 1.70 | -1.83 |
| 3 | 265.3 | -5.3 | 1.53 | -2.03 |
| 4 | 275.0 | -3.7 | 1.63 | -1.36 |
| 1976:1 | 284.4 | -2.2 | 2.04 | -0.79 |
| 2 | 294.0 | -5.6 | 2.58 | -1.96 |
| 3 | 308.2 | -3.9 | 2.87 | -1.30 |
| 4 | 321.7 | -1.5 | 3.08 | -0.48 |

RMS = 3.85 $U^2$ = 0.170

TABLE 7

RATE OF CHANGE-- N-PERIODS (N=3)

| | $Y_p$ | $Y_p-Y_a$ | $P_i$ | $P_i-A_i$ |
|---|---|---|---|---|
| 1975:1 | 254.3 | 1.9 | 2.00 | 0.76 |
| 2 | 256.8 | -4.5 | 1.74 | -1.79 |
| 3 | 266.5 | -4.1 | 1.99 | -1.57 |
| 4 | 277.4 | -1.3 | 2.51 | -0.48 |
| 1976:1 | 287.1 | 0.5 | 3.01 | 0.18 |
| 2 | 296.6 | -3.0 | 3.49 | -1.05 |
| 3 | 310.8 | -1.3 | 3.74 | -0.43 |
| 4 | 323.8 | 0.6 | 3.75 | 0.19 |

RMS = 2.59 $U^2$ = 0.083

The high degree of accuracy achieved by these relatively simple techniques as applied to this series indicates the existence of a definite, persistent trend in the series. It is in a series of this nature that simple trend projection techniques such as these are most applicable. A trend projection from one period is incapable of forecasting any turning point in a series, and such a forecast is very improbable, although possible, with an n-period trend projection method. An n-period projection method could forecast a return to a long-run trend after a short deviation from such a trend, such as a forecast of an increase in a series with long-run growth after a one period decline. However, the presence of turning points would be likely to affect the accuracy of the forecast.

These trend projection methods would not be expected to be very accurate in forecasting a highly volatile series with frequent turning points. One period trend projections would be accurate only in the absence of any turning points, and therefore unreliable to the extent that such turning points are probable in the series. N-period projection methods would be only slightly more reliable, their accuracy affected by off-setting increases and decreases in the series.

## D. Seasonal Methods

If a series shows regular variation corresponding to certain seasons of the year, this seasonality must be taken into consideration in making a forecast. Such series with seasonality would be typified by the increase in toy sales around Christmas or the decrease in construction activity during the winter months. These series cannot be accurately forecast without taking their seasonality into account.

The simplest forecast method incorporating seasonality is a seasonal continuity method. This method modifies the random walk process by assuming that the change in the value of the variable from one period to the same season of the next year follows a random walk process. Thus, a series of quarterly data is assumed to be composed of four different random walks, one for each quarter, and the forecast is given by

$$Y_{t+1} = Y_{t-3}$$

For monthly data, the forecast would be

$$Y_{t+1} = Y_{t-11}$$

Forecasts for the mortgage series using this method are given in Table 8.

TABLE 8

SEASONAL CONTINUITY

| | $Y_p$ | $Y_p - Y_a$ | $P_i$ | $P_i - A_i$ |
|---|---|---|---|---|
| 1975:1 | 236.5 | -15.9 | -5.13 | - 6.37 |
| 2 | 243.8 | -17.5 | -3.41 | - 6.94 |
| 3 | 247.6 | -23.0 | -5.24 | - 8.80 |
| 4 | 249.3 | -21.3 | -7.87 | -10.86 |
| 1976:1 | 252.4 | -34.2 | -9.44 | -12.27 |
| 2 | 261.3 | -38.3 | -8.83 | -13.37 |
| 3 | 270.6 | -41.5 | -9.68 | -13.85 |
| 4 | 278.7 | -44.5 | -10.70 | -14.26 |

RMS = 31.39 $U^2$ = 10.696

These results indicate the lack of any significant seasonality in the mortgage series. The error in these predictions, greater than those of any method yet examined, shows that the method is not applicable to this series.

Seasonality may also be examined, and used for forecasting, through the use of a seasonal index. The first step is to calculate a moving average for the series; a four-quarter moving average for quarterly data, a twelve month moving average for monthly data. The average value corresponding to each period is then divided into the actual value for that period, and the result multiplied by 100. The resulting index numbers for a period of years are then averaged by quarter or by month to obtain the average index number for each period.

Any series measured in increments of time less than one year is made up of four components which determine the series. These four components are trend, seasonal, cyclical, and irregular. These may be represented by T,S,C, and I respectively. Averaging the series over a year eliminates the seasonal component, 'S'. The moving average then contains trend, cyclical, and irregular components. Dividing this average back into the original data isolates the seasonal component. Assuming that the components interact in a multiplicative manner,

$$TSCI \; / \; TCI = S.$$

This result, times 100, gives an index of seasonality.

This procedure was applied to the mortgage series using five years of quarterly data (1970:1 to 1974:4). The index numbers are given in Table 9. If the series had shown strong seasonality, the forecast value for a given season would be adjusted by that index number.

TABLE 9

SEASONAL INDEX

| Quarter | Index |
|---|---|
| 1 | 99.3 |
| 2 | 100.0 |
| 3 | 100.5 |
| 4 | 100.2 |

Calculation of seasonal indexes confirms the absence of any noticeable seasonality in the mortgage series, at least in recent years. Seasonal forecasting techniques are best applied to a series with regular seasonal fluctuations around a fairly stable mean. If the series shows long-run movement in one direction as well as seasonality, a combination of a trend projection method adjusted by a seasonal technique would be expected to give more accurate forecasts than would either method by itself.

# CHAPTER III

## INDICATORS

### A. Leading Series

One of the basic problems found with the use of naive forecasting techniques, particularly simple trend projection methods, is that such methods are generally incapable of forecasting turning points in a series. In working with a highly volatile series, accuracy in prediction of turning points may be more important than an accurate prediction of the magnitude of change. A simple means of predicting change in direction for a series is through the use of indicators, or lead-lag series. The major work in this area is done by the National Bureau of Economic Research, and published in Business Conditions Digest by the U. S. Department of Commerce.

Study of major series of economic variables in relation to the general business cycle has lead to the classification of these series into three major categories: leading indicators or series, roughly coincident series, and lagging series. Classification is made according to the timing relationship of peaks and troughs of the individual series to peaks and troughs in the general business cycle.

> Two lists of economic indicators are regularly available for use by the analyst. The full list includes 88 U. S. series: 36 leading, 25 roughly coincident, and 16 unclassified by timing. Of these series, 72 are monthly and 16 quarterly. The short list includes 25 U.S. series: 12 leading, 7 coincident, and 6 lagging. There is little economic sector duplication on this list of 21 monthly and 4 quarterly series, all of which are included in the long list.[1]

Leading indicators include such series as the average work week (manufacturing) and an index of common stock prices. Among the roughly coincident series are the unemployment rate and personal income. Lagging series include interest rates and manufacturers' inventories.

Although the turning point timing of these series is not precise, the average lead time of one or more leading series can be very useful in predicting turning points in other series or in the general business cycle. This can be done by way of a subjective evaluation of one or more series which the forecaster considers to be relevant to his area of concern, or through the construction of one or more indexes based on such series. The methods of constructing such indexes will be discussed below.

As an aid in evaluating the strength and reliability of various series, the National Bureau has assigned scores to each series.

> For the recent revision, a scoring system was developed to aid in the selection process. It does not eliminate the need for decision and choice,

---

[1]R.K.Chisholm and G.R.Whitaker, Forecasting Methods (Homewood: Irwin, 1971), p.43.

> but it serves as a rough guide. Further, it is a convenient way for the researchers to account for the factors considered and the weighting scheme used in making their choice...Of the many time series considered, 120 were assigned scores. On the basis of 100, the average score was 62, the range 35-89. [1]

Each series is scored via a weighted consideration of six major characteristics. These are economic significance, statistical adequacy, conformity (to past cycles), timing (lack of large variability), currency (when data is available), and smoothness.

Examination of the mortgage series indicates that the use of indicators with this series might be impractical. This is because the series does not have any turning points, but exhibits persistent growth over time. Quarterly data from 1960 to the present shows not one turning point.

However, there are variations in the rate of growth of the series. If first differences of the series are taken, and the resulting series of amount of change in mortgage debt examined, definite turning points do exist. This new series also shows strong seasonality, perhaps due to the influence of some factor or factors which affect the rate of growth but are not strong enough to affect the absolute level. The first-differenced series was adjusted via a moving average (monthly data used to correspond to National Bureau series) and compared with the turning points in general economic activity. The results are as follows for the most recent turning points.

---

[1] Chisholm and Whitaker, op. cit., pp.46-47.

TABLE 10

TURNING POINTS

| | Gen. | Mort. | Lead |
|---|---|---|---|
| Peak | Nov. '69 | Feb. '69 | - 9 |
| Trough | Nov. '70 | Jan. '70 | -10 |
| Peak | Nov. '73 | Dec. '72 | -11 |
| Trough | Mar. '75 | Sep. '74 | - 6 |

It thus appears that the series of changes in mortgage debt is a leading series, with turning points occurring significantly in advance of turning points in the general level of economic activity. With the above indication that the series leads general activity by an average of nine months, it would seem that the method of forecasting with leading indicators is not applicable to the mortgage series, since the National Bureau leading series in general lead by 2 to 10 months. The series can, however, be used as an indicator of turning points for other series. In addition to the long lead time shown by this series, there is also the fact that the first-differenced series showed no other turning points over the period examined except those four corresponding to those of the general cycle.

B. Diffusion Index

A simple method of quantifying the activity of a number of series is through the use of a diffusion index. This index simply states what percent of a given number of series under consideration are increaaing at a given time. It is calculated by dividing the number of rising series by the total number of series in the index and multiplying by 100. An index value above 50 indicates that the aggregate series is rising, and below 50 that it is declining. Thus, a diffusion index of leading indicators falling below 50 is taken as an indication of a future downward turning point in the general level of economic activity. In general, a diffusion index does not indicate magnitude of change, nor does it give different weights to the various series which go to make it up.

The National Bureau computes several indexes for different series, which are published in Business Conditions Digest.

> Diffusion indexes based on these series behaved perfectly in postwar cycles....The diffusion indexes, along with the leading indicators, have indicated every turn in general economic activity. They have also turned for a short period when no turn occurred, however, so they cannot be used alone as a forecasting device or without judgement. They also do not tell the exact timing of turning points or the intensity of changes in economic activity.[1]

## C. Pressure Indexes

A pressure index is an attempt to make use of the magnitudes of various series to aid in forecasting. This involves the use of various ratio and difference measures of two important series to predict changes in a third. For example, the ratio of raw material inventories to new orders for finished goods in an industry may be useful in forecasting changes in raw material prices. As another

---

[1] Carl A. Dauten and Lloyd M. Valentine, Business Cycles and Forecasting (Cincinnati: South-Western, 1974), pp.336-337.

example, the ratio of railroad car loadings to the number of cars in service may indicate the demand for new cars. Specialized knowledge of a particular industry and of relevant economic series aid in forecasting via pressure indexes.

A pressure index is aimed at forecasting turning points rather than specific values.

> These ratio and difference measures...may not always be helpful in forecasting the magnitude of change. However, they do serve the useful purpose of providing warning signals of impending developments, and frequently an indication as to the future of change. When used in conjunction with other forecasting methods, pressure indexes can accomplish much in the way of establishing guideposts for future planning.[1]

Again, because of the nature of the mortgage series, it was necessary to first difference the series in applying a pressure index technique. Testing and examining various series, it was found that the supply side of the mortgage market seemed to be more stable in its relationship to the series, and therefore more useful for forecasting.

Assuming that more funds would be available for mortgages as the return on these funds becomes more attractive, a comparison was made between mortgage interest rates (conventional first mortgages on new homes) and alternative investment opportunities, as represented by the return on prime commercial paper. Dividing the return on mortgages by the return on commercial paper gives a ratio which

---

[1]Spencer and Siegelman, op. cit., p.35.

might indicate the relative attractiveness of investment in the mortgage market, and therefore an increase in the amount of funds supplied to the market as thus ratio increases. Listed below are the most recent turning points in the ratio and in the first differenced mortgage series.

TABLE 11

PRESSURE INDEX

| | Ratio | Mort. | Lead |
|---|---|---|---|
| Peak | Oct. 68 | Feb. 69 | - 4 |
| Trough | Jul. 69 | Jan. 70 | - 6 |
| Peak | Feb. 72 | Dec. 72 | -10 |
| Trough | Jul. 74 | Sep. 74 | - 2 |

While these results are inconclusive and vary widely, they do indicate that the interest ratio may give some useful information about future changes in the mortgage series.

D. Amplitude-Adjusted Index

An amplitude-adjusted index is an index constructed from a number of series, taking into account the rate of change in these series. As with other indexes, a number of series are included on the basis of relevance to the variable being forecast.

For each included series, the month-to-month percent change is computed and the resulting percentage series is standardized about a mean of one. Then for each month a weighted average of these standardized values is computed. In computing this average, the weight for each series is the National Bureau of Economic Research score for that series. The new time series of weighted averages is then standardized about a mean of one.

> Although this new series is a composite of the several series involved, the index number itself has an interpretation which the forecaster may find useful. If the value of the index for the most recent month is greater than one, the various time series are rising together more rapidly than they have on average in the past. This would be taken to mean that a relatively strong expansion is underway. If, however, the value is less than one, the series are rising less rapidly than they have on average in the past. This would be taken to mean that the expansion is relatively weak or about to reverse. [1]

An index of this type seems to be better suited for forecasting changes in aggregate activity than for individual industry. Measurement of the movement of a number of economic series would lack the accuracy available from other techniques in forecasting one particular economic variable.

Both the construction and interpretation of this and other indexes can be a very subjective matter, relying on specific knowledge of the area involved. While naive methods are of little value in forecasting turning points, indicators and indexes give little or no information about magnitude of change. Perhaps a forecast derived from the use of a combination of the two methods could be used with more confidence than either method by itself.

---

[1] Chisholm and Whitaker, op. cit., p.50.

# CHAPTER IV

## REGRESSION ANALYSIS

### A. The Use of Regression

Regression analysis is the statistical analysis of the relationship between the values of two or more variables. The variables to be analyzed are generally chosen because of theoretical relationships between them.

> Probably the chief purpose of regression analysis is to predict the value of a dependent variable from firm estimates or values of independent variables which are highly correlated with the dependent variable.[1]

By use of the historical data of the variables, regression analysis is used to derive a mathematical statement of the relationships between the variables.

> (An) equation...expresses the relationship between a dependent and one or more independent variables. ...in a single-equation model it is assumed that the relationship is...one way in nature so that variations in the independent variables cause variations in the dependent variable, but not vice versa...[2]

Once the historical relationship between variables has been quantified, it may be used to forecast future values for the dependent variable. If the values of the independent variables are known (as in the case of lagged

---

[1]Howard Balsley, Quantitative Research Methods for Business and Economics (New York: Random House, 1970), p.183.

[2]Spencer and Siegelman, op. cit., p.44

variables) or can themselves be accurately forecast, then a forecast for the dependent variable can be made from these values together with the mathematical expression of the relationship. Thus, regression analysis may be used to both analyze historical data and to predict future values of the variable under consideration.

## B. The Model

The basic regression model is the ordinary least squares (OLS) regression analysis. This model or method computes a relationship between variables (i.e., fits a line to plotted values of the independent and dependent variables) such that the sum of the squared residuals (deviations from the line) is less than for any other line. The use of squared deviations eliminates the tendency of positive and negative deviations to cancel out.

A simple model of the relationship between two variables, X and Y, is given by the expression

$$Y = a + bX$$

where 'a' and 'b' are constants computed by the regression analysis. The constant 'a' represents the Y-axis intercept and 'b' is the slope of the regression line.

The relationship described by the above equation is a simple linear relationship between two variables. It may be that the relationship is better described by a non-linear equation, such as

$$Y = aX - bX^2$$

which is a parabolic curve, or any one of many possible curvilinear relationships.

However, the assumption is generally made that the relationship is a linear one. As Dutta says,

> Econometric research is largely restricted to using linear functions or nonlinear functions which can be appropriately transformed to linear ones. The assumption of a linear relationship is for convenience.[1]

The assumption of linearity is made to avoid the complex computations needed to calculate curvilinear relationships. If a nonlinear function must be used, a specification with linear parameters, or constant terms, such as

$$Y = a + bX + cX^2$$

is generally preferred to one with exponential parameters, where the constant appears as an exponent, such as

$$Y = aX^b$$

Again, the simplified calculation involved is considered to be worth any goodness-of-fit sacrificed because of the linear assumption. As Klein says,

> ...we usually go to great lengths to keep the forlations of demand function linear in the parameters, whereas we frequently drop the linearity assumption for the variables.[2]

One method of eliminating exponential parameters is through the use of logs. Since the term $X^b$ may be expressed as $b \ln X$, a regression of a nonlinear function may be computed by the linear format

$$\ln Y = a + b \ln X$$

---

[1] Dutta, op. cit., p.15

[2] Lawrence Klein, An Introduction to Econometrics (Engelwood Cliffs: Prentice-Hall, 1962), p.23.

## C. Tests of Fit

Various statistical teats exist to compare the results of different regressions in order to see which of several computed lines best fit the data for which they are computed.

The coefficient of determination ($R^2$) is an estimate of the amount of the total variation in the dependent variable which is explained by the regression model. Variation in the dependent variable (Y) can be broken down into

1. Total variation: $SS_t = \sum (Y - \bar{Y})^2$
2. Explained var.: $SS_e = \sum (Y_c - \bar{Y})^2$
3. Unexplained var.: $SS_u = \sum (Y - Y_c)^2$

where Y is the observed value of the dependent variable, $Y_c$ is the value computed from the regression equation, and $\bar{Y}$ is the mean of the observations. $R^2$, the proportion of the total variation which is explained by the model, is computed by the expression

$$R^2 = SS_e / SS_t = 1 - (SS_u / SS_t)$$

$R^2$ has a range from zero, indicating that there is no correlation between changes in the independent variable and changes in the dependent variable, to one, indicating perfect correlation between changes in the variables.

The F-statistic is a test of significance of the computed equation. It is used to test whether or not the regression coefficients are dignificantly different from zero. The statistic is calculated as

$$F = SS_e / SS_u$$

and used to test the null hypothesis that all the regression coefficients are equal to zero. The computed value of 'F' is compared with a table value for the appropriate number of degrees of freedom (the number of changeable variables subject to the constraints of the formulation). If the computed value is greater than the table value, the regression coefficients are taken to be significantly different from zero, and the null hypothesis is rejected. As the computed value of 'F' increases, more of the variation in the dependent variable is explained by the regression, since 'F' increases as $SS_e$, the explained variation in Y, increases.

The unexplained variation in the dependent variable is often referred to as the error term, and may be designated as 'u'. While the regression equation, and each calculated value of the dependent variable, is expressed as

$$Y = a + bX$$

the observed value in these terms is expressed as

$$Y = a + bX + u.$$

The error term, or disturbance, is that part of the variation in the dependent variable which is not systematic and cannot be estimated. The presence of an unknown factor can make forecasting difficult. Dutta discusses the problem as follows.

> One possibility is to make some assumptions on the nature of the probability distribution of the random disturbance term U which we cannot observe.... In the absence of any knowledge of the nature of the distribution, it will usually be assumed that the disturbances U are random variables with zero mean

> and constant and finite variance, and that they are drawn independently from this unknown distribution.[1]

Some other tests of the goodness-of-fit of the regression equation include the standard deviation of the error term and the average absolute error.

> While the coefficient of determination is often a convenient figure of merit in assessing the goodness-of-fit of a particular equation to a set of data, occasions arise where an even more direct measure of explained and unexplained variation can be helpful. The standard deviation of the error term u is such a measure, as is the average absolute error.[2]

The standard deviation of the error term is the root mean square value of the error term. It is computed as

$$\sigma_u = \sqrt{\frac{1}{n}\sum u^2}$$

> Defined this way, (it) can be interpreted as a measure of the effective standard deviation of Y, once knowledge of the independent variables and the equation parameters has been used to compute $Y_c$, i.e., once we have explained a portion of the inherent variation in Y through the use of the regression equation.[3]

Since the true values of 'u' depend on knowledge of the true parameters of the relationship, only an estimate of the standard deviation of the error term can be made ($S_u$). This is given by

$$S_u = \sqrt{\frac{1}{n-2}\sum u^{*2}}$$

where u* is the calculated error term from the estimated regression model.

---

[1]Dutta, op. cit., p.27.

[2]Elliott, op. cit., pp.37-38.

[3]Ibid., p.38

A measure of the average amount of error is given by the mean absolute error.

$$E = \frac{1}{n} \sum \left| Y - Y_c \right|$$

This measurement allows a better conceptualization of the amount of error in an equation since it does not give added weight to extreme errors by squaring them, as does the standard deviation.

It is also possible to test the significance of the individual parameters of a regression equation, in order to see if they are significantly different from zero.

> ...the partial nature of regression coefficients means that the quantitative effect of one influential variable upon Y can be isolated from the effects of others.[1]

The partial effect of one independent variable on the dependent variable can be tested via a significance test of the individual coefficients. This is done with a t-test. The t-statistic is computed as

$$t = b/S_b$$

where 'b' is the computed parameter and $S_b$ its estimated standard deviation. Then is then used to test the null hypothesis that the parameter 'b' is equal to zero. If the computed value of 't' is greater than the appropriate table value, the null hypothesis is rejected and the parameter is taken to be significantly different from zero.

## D. Statistical Problems

Various statistical problems can affect the reliability

---

[1] Elliott, op. cit., p.42

of the OLS regression analysis. These problems involve violations of the basic assumptions about the nature of the error term of the regression model. These assumptions are that the error term is a random variable with zero mean and constant and finite variance. Among these problems, some of the more important are autocorrelation, heteroscedasticity, and multicollinearity.

### D1. Autocorrelation

Autocorrelation, or serial correlation, is a violation of the assumption that the error terms are random and statistically independent of one another. If the error terms are serially correlated, knowledge of the value of $u_{t-1}$ improves the ability to predict the value of $u_t$. This would not be true if the error terms were random and independently drawn from this distribution. Positive autocorrelation occurs when the occurance of an error term raises the probability above 0.5 that the next error term will have the same sign. Negative autocorrelation is the presence of sequential error terms alternating in sign. Either case may be an indication that an important independent variable has been omitted in the specification of the regression equation.

A regression specification may be tested for autocorrelation by inspection of the residuals for sign patterns, either alternating signs throughout or long runs of the same sign, or by use of the Durbin-Watson statistic. In the presence of autocorrelation, conclusions based on

$R^2$ statistic and t-tests are likely to be unreliable. In particular, a high value for $R^2$ may be unwarrented. A regression model may be estimated in the presence of autocorrelation through the use of a correction procedure, such as the first-difference method. This involves using the amount of change in each variable for each period, rather than using the original data, and running a regression on these series of changes.

Autocorrelation is generally present to some degree in time series data. The long-term growth inherent in most time series of economic variables makes autocorrelation a persistent problem. Because of the upward trend over-time, the series are not truly random. The seriousness of the problem depends upon the purpose of the regression analysis. If the purpose is to determine the underlying relationships between economic variables, autocorrelation makes this difficult, particularly if it has been caused by the omission of an important variable.

> ...whenever truly influential variables have been omitted from a regression equation, their impact will be taken up by included variables to the extent of the correlation between those variables included and those excluded. This will influence the magnitudes of included regression coefficients...[1]

Thus, estimation of economic relationships in the presence of autocorrelation can lead to attaching too much importance to the role of one independent variable in determining the dependent variable, at the cost of underplaying

[1]Elliott, op. cit., p.63.

or ignoring one or more other variables.

For the purpose of forecasting, autocorrelation may not present a serious problem. If the effect of an omitted variable is consistently and accurately taken up by the included variables, forecasts made with the model will be acceptable. The usefulness of a forecast comes from its accuracy, not necessarily from its ability to describe the underlying economic relationships. If a model displays serial correlation, other variables may be included in the specification or corrections such as the first-difference method tried, the most accurate result then being adopted for forecasting purposes.

## D2. Heteroscedasticity

Heteroscedasticity is a violation of the assumption of constant error variance, where the variance of the error term is not constant in size across the observations. In general, the variance of the error term increases either directly or inversely with changes in the value of the independent variable. This may indicate correlation between the error term and the independent variable. Heteroscedasticity is more likely to occur in cross-sectional data than in time series data. For example, consumption spending is likely to be more volatile in upper-income groups. Its presence in time-series data may indicate a change in the structure underlying the observations, such as a change in government policy during the observation period.

The presence of heteroscedasticity distorts the measurement of the error terms, and therefore the reliability of significance tests such as $R^2$. One method of checking for heteroscedasticity is the Goldfeld-Quandt Test. This involves computing on regression for that portion of the data assumed to have low variance of the error term, and another regression for the remainder of the data. The residual variations are then compared to see if they are significantly different from one another. Another method is to rank order the independent variables and inspect the associated error terms. Heteroscedasticity shows up as a systematic relation between the absolute value of the error term and the size of the independent variable.

There are various ways to correct for heteroscedasticity. One way is to deflate all variables by a common term which is linearly related to the error term. In the case of heteroscedasticity caused by changes in the underlying structure or relationships, the problem may be eliminated by using only those observations occurring since the estimated time of the change.

### D3. Multicollinearity

In multiple regression analysis, where two or more independent variables are used in the equation, it is assumed that the effects of these variables are independent, linear, and additive. Multicollinearity is a violation of this assumption; it is the presence of a high degree of correlation between two or more independent variables. In the presence of multicollinearity, the regression coefficients may be unreliable and volatile, with the addition or omission of a few observation greatly changing the estimates of the parameters.

Multicollinearity may be indicated by a high partial correlation between independent variables or by high standard errors for the regression parameters.

> If several coefficients have high standard errors and dropping one or more variables from the equation lowers the standard errors of the remaining variables, multicollinearity will usually be the source of the problem.[1]

If the model is to be used to analyze theoretical relationships between economic variables, the presence of multicollinearity can be quite serious, as Elliott points out.

> As multicollinearity damages the reliability of obtained regression coefficients, it also damages the ability to draw well-supported conclusions about the significance of individual variables from t-tests. We cannot reliably make inferences about the statistical importance of individual variables under these circumstances.[2]

However, if the model is to be used only for forecasting purposes, multicollinearity may present little or no problem.

> If predicting the dependent variable is the sole purpose of the model and if whatever multicollinear pattern present in the data persists into the forecast period, then the equation may produce reasonable forecasts in spite of the somewhat arbitrary weighting attached to individual intercorrelated variables.[3]

---

[1]Robert Pindyck and Daniel Rubinfeld, Econometric Models and Economic Forecasts (New York, McGraw-Hill, 1976), p.68

[2]Elliott, op. cit., p.60.

[3]Ibid., p.61

The forecaster should be aware of the presence of multicollinearity, through the inspection of standard errors and partial correlations. Any changes in the interdependence of the independent variables must be noted and taken into account in the specification of the model.

As may be noted, multicollinearity shows a structure somewhat similar to that of autocorrelation. While the former is an interdependence of independent variables, the latter is an interdependence of error terms. The first may be corrected by dropping a variable, and the second by adding one.

# CHAPTER V

# FORECASTING WITH REGRESSION ANALYSIS

## A. Time Series Regression

Time series regression is a method of trend projection, along the lines of the naive methods discussed earlier, using the mathematics of regression analysis. Here, OLS regression is used to compute the line of best fit between a dependent variable and time. This method assumes that the series moves in a particular direction over time and that the underlying factors determining this movement are stable enough and consistent enough to be projected into the future.

The time series model takes the form

$$Y = a + bT$$

where Y is the dependent variable and T is the time period.

A regression of the mortgage series against time, using 20 observations, 1970:1 to 1974:4, yields the model

$$Y = \underset{(1.893)}{126.9} + \underset{(0.158)}{6.425}T \qquad R^2 = .9892$$

The numbers in parenthesis are the standard errors of the computed coefficients. Forecasts based on this model are given in Table 12.

TABLE 12

TIME SERIES (N=20)

| | $Y_p$ | $Y_p - Y_a$ | $P_i$ | $P_i - A_i$ |
|---|---|---|---|---|
| 1975:1 | 261.8 | 9.4 | 5.01 | 3.77 |
| 2 | 268.2 | 6.9 | 6.26 | 2.72 |
| 3 | 274.7 | 4.1 | 5.13 | 1.57 |
| 4 | 281.1 | 2.4 | 3.88 | 0.89 |
| 1976:1 | 287.5 | 0.9 | 3.16 | 0.33 |
| 2 | 293.5 | -5.7 | 2.56 | -1.98 |
| 3 | 300.4 | -11.7 | 0.28 | -3.89 |
| 4 | 306.8 | -16.3 | -1.70 | -5.26 |

RMS = 8.62 $U^2$ = 0.761

It is possible that the most recent observation of a series might provide the best information upon which to base forecasts. The next group of forecasts were made from a series of regressions using constantly updated data. For each quarterly forecast, a new regression was computed based on the eight preceding quarters.

TABLE 13

UPDATED TIME SERIES (N = 8)

| | $Y_p$ | $Y_p - Y_a$ | $P_i$ | $P_i - A_i$ |
|---|---|---|---|---|
| 1975:1 | 257.1 | 4.7 | 3.13 | 1.89 |
| 2 | 258.6 | -2.7 | 2.46 | -1.07 |
| 3 | 263.7 | -6.9 | 0.92 | -2.64 |
| 4 | 271.8 | -6.9 | 0.44 | -2.55 |
| 1976:1 | 284.9 | -1.7 | 2.22 | -0.61 |
| 2 | 289.6 | -10.0 | 1.05 | -3.49 |
| 3 | 302.5 | -9.6 | 0.97 | -3.20 |
| 4 | 317.0 | -6.2 | 1.57 | -1.99 |

RMS = 6.69 $U^2$ = 0.477

That these forecasts are more accurate than those given by the simple time-series regression analysis could indicate that while the mortgage series exhibits continued growth, its rate of growth might not be constant. If this is the case, for example if it is increasing at an increasing rate, a nonlinear function might give a regression line with better fit and more accurate forecasts. A linear regression function assumes a constant rate of change, which a curvilinear model does not. The forecasts in Table 14 were derived from the model

$$Y = 128.1 + 6.1T + 0.016T^2 \qquad R^2 = .988$$

TABLE 14

CURVILINEAR TIME SERIES

| | $Y_p$ | $Y_p - Y_a$ | $P_i$ | $P_i - A_i$ |
|---|---|---|---|---|
| 1975:1 | 263.3 | 10.9 | 5.62 | 4.38 |
| 2 | 270.0 | 8.7 | 6.97 | 3.44 |
| 3 | 276.9 | 6.3 | 5.95 | 2.41 |
| 4 | 283.7 | 5.0 | 4.84 | 1.85 |
| 1976:1 | 290.6 | 4.0 | 4.27 | 1.44 |
| 2 | 297.5 | -2.1 | 3.80 | -0.74 |
| 3 | 304.5 | -7.6 | 1.64 | -2.53 |
| 4 | 311.4 | -11.8 | -0.22 | -3.78 |

RMS = 7.72 $\quad U^2 = 0.674$

The curvilinear trend model doesn't give noticeably better results than the updated method, although both are more accurate than the simple trend regression. This would indicate the possibility that the basic structure of the series is changing, or that it is curvilinear.

The fact that these regression techniques do not give forecasts as accurate as those of simple naive trend projection methods discussed previously leads toward the conclusion that it is the most recent observations of this series which may have the most significance for time-series analysis and that inclusion of an excessive number of earlier observations may lead to a decrease in accuracy. The time period used in estimating the model can be varied in testing and the most accurate result used for forecasting.

As with simple trend projection techniques, time-series regression provides more accurate forecasts for series which show long-run, consistent movement in one direction. A stable series, free of wide fluctuations, may be much more accurately forecast by these methods than a series with frequent turning points. Just as with simple trend projection methods, it is not likely that any turning point in a series would be forecast with a time series regression model.

## B. Autoregression

Just as a series may be regressed against time in order to calculate a trend line, so it may be regressed against lagged values of itself. A regression of the variable against lagged values of the series will indicate the presence of any systematic pattern of movement in the series, any consistent growth patterns.

An estimation of the model

$$Y_{t+1} = a + bY_t$$

for the mortgage series yields the equation

$$Y_{t+1} = \underset{(1.133)}{-1.5142} + \underset{(.007)}{1.0331}Y_t \qquad R^2 = .9982$$

Forecasts based on this equation are given in Table 15.

TABLE 15

AUTOREGRESSION (N=40)

| | $Y_p$ | $Y_p-Y_a$ | $P_i$ | $P_i-A_i$ |
|---|---|---|---|---|
| 1975:1 | 256.0 | 3.6 | 2.69 | 1.45 |
| 2 | 259.2 | -2.1 | 2.69 | -0.84 |
| 3 | 268.4 | -2.2 | 2.72 | -0.84 |
| 4 | 278.0 | -0.7 | 2.73 | -0.26 |
| 1976:1 | 286.4 | -0.2 | 2.76 | -0.07 |
| 2 | 294.6 | -5.0 | 2.79 | -1.75 |
| 3 | 308.0 | -4.1 | 2.80 | -1.37 |
| 4 | 320.4 | -2.3 | 2.82 | -0.74 |

RMS = 2.95 $\quad U^2 = 0.096$

If a time series appears to follow a nonlinear trend, such as one which is increasing at an increasing rate, a nonlinear autoregression may be estimated. In order to avoid excessively complicted calculations, regression models are usually specified in such a manner as to avoid exponential parameters. One method of doing this is through the use of logrithms. A log form autoregression model is

$$\ln Y_{t+1} = a + b \ln Y_t$$

Estimating for the mortgage series,

$$\ln Y_{t+1} = \underset{(0.082)}{0.062} + \underset{(0.016)}{0.994} \ln Y_t \qquad R^2 = .9956$$

The antilog of $\ln Y_{t+1}$ gives the forecast for the period t+1. Forecasts based on this model are in Table 16.

TABLE 16

LOG AUTOREGRESSION (N=20)

| | $Y_p$ | $Y_p-Y_a$ | $P_i$ | $P_i-A_i$ |
|---|---|---|---|---|
| 1975:1 | 256.6 | 4.2 | 2.93 | 1.69 |
| 2 | 259.8 | -1.5 | 2.93 | -0.60 |
| 3 | 268.9 | -1.7 | 2.91 | -0.65 |
| 4 | 278.4 | -0.3 | 2.88 | -0.11 |
| 1976:1 | 286.7 | 0.1 | 2.87 | 0.04 |
| 2 | 294.8 | -4.8 | 2.86 | -1.68 |
| 3 | 308.4 | -3.7 | 2.94 | -1.23 |
| 4 | 320.8 | -2.4 | 2.79 | -0.77 |

RMS = 2.85 $U^2$ = 0.091

Both autoregressive methods give better results than any of the time-series regressions discussed in the previous section, and are in the same accuracy range of the simple trend projection techniques covered earlier. Like all trend projection techniques, autoregression methods lack the ability to accurately forecast deviations from the long-run trend of a series.

In forecasting a series with persistent growth such as the mortgage series, forecasts will generally be accurate as long as the growth pattern continues. Any turning point in the series will generally be missed using these techniques, and subsequent forecasts will be unreliable as long as the series continues in a downward movement.

Again, autoregressive and time-series regression methods work best in forecasting series with consistent, long-run movement in one direction, and are generally not reliable in predicting a series showing frequent turning points. In a series showing only occasional turning points followed by a number of periods of movement in one direction, these methods should give fairly accurate forecasts for a majority of the periods. They would however miss the critical turning points.

## C. Simple Regression Models

Forecasting with models of this type is based upon correlation between two series of economic variables. The relationship between the variable to be forecast and another economic variable which is thought to influence or determine the first is computed via the regression method and used to forecast future values for the dependent variable.

It may be that the variable under consideration does not fit neatly into the conventional terms of economic theory. For example, the mortgage series represents the outstanding mortgage debt held by saving and loan associations. If the mortgage market is in equilibrium, this series should then represent both the supply of mortgage funds and the demand for mortgage loans at the point of equilibrium. However, if the mortgage market is not in equilibrium, the series could represent either supply or demand, or something somewhere in between the two. As with other problems considered, this question of identification may not be of serious concern to the forecaster. The forecaster is concerned with establishing a mathematical relationship which gives accurate forecasts for the dependent variable. He is not directly concerned with whether or not this corresponds exactly to the true underlying economic relationships between variables.

If the mortgage series represents demand, it could be theorized that it is a function of personal income. As personal income increases, more people would want to buy homes and thus the demand for mortgage credit would increase. A regression of the mortgage series with personal income gives the equation

$$Y = \underset{(5.37)}{-29.556} + \underset{(.007)}{0.233PI} \qquad R^2 = .97$$

This and all subsequent regressions were run on ten years of quarterly data, 1965:1 to 1974:4, unless otherwise noted.

TABLE 17

PERSONAL INCOME

| | $Y_p$ | $Y_p-Y_a$ | $P_i$ | $P_i-A_i$ |
|---|---|---|---|---|
| 1975:1 | 248.8 | - 3.6 | -0.20 | -1.44 |
| 2 | 260.3 | - 1.0 | 3.13 | -0.40 |
| 3 | 268.0 | - 2.6 | 2.56 | -1.00 |
| 4 | 275.2 | - 3.5 | 1.70 | -1.29 |
| 1976:1 | 283.1 | - 3.5 | 1.58 | -1.25 |
| 2 | 289.1 | - 9.9 | 1.08 | -3.46 |
| 3 | 294.7 | -17.4 | -1.64 | -5.81 |
| 4 | 306.1 | -17.1 | -1.92 | -5.48 |

RMS = 9.61 $U^2$ = 0.872

From the other side of the market, the supply of mortgage funds could be a function of the total assets of saving and loan associations. If saving and loans hold a stable proportion of their assets as mortgage debt, this could give an indication of the supply of funds to the market. A regression between these two variables gives the equation

$$Y = \underset{(.724)}{2.285} + \underset{(.004)}{0.837}A \qquad R^2 = .9992$$

where Y is mortgage debt and A is total assets.

TABLE 18

ASSETS

| | $Y_p$ | $Y_p-Y_a$ | $P_i$ | $P_i-A_i$ |
|---|---|---|---|---|
| 1975:1 | 257.5 | 5.1 | 3.29 | 2.05 |
| 2 | 268.4 | 7.1 | 6.34 | 2.81 |
| 3 | 277.7 | 7.1 | 6.28 | 2.72 |
| 4 | 285.5 | 6.8 | 5.51 | 2.52 |
| 1976:1 | 298.3 | 11.7 | 7.03 | 4.20 |
| 2 | 309.1 | 9.5 | 7.85 | 3.31 |
| 3 | 320.3 | 8.2 | 6.91 | 2.74 |
| 4 | 329.6 | 6.4 | 5.61 | 2.05 |

RMS = 7.97 $\qquad U^2 = 0.702$

The regression on assets gives accuracy test results which are a little better than the regression on income. However, these results do not compare favorably with those of some methods discussed earlier, especially the naive trend projection methods. It appears that forecasts based on regression techniques need further work to improve their accuracy in forecasting this series.

## D. Lagged Variables

The above forecasts were made on the basis of correlation of the mortgage series with values of the independent variables for the same period. This presents no problem in testing a method via retrospective "forecasts", as the values of the independent variables are already known. However, for purposes of making actual forecasts, it is necessary to also forecast the values of the independent variables for the forecast period. This problem can be avoided through the use of lagged variables. Here, the dependent variable is correlated with the value of the independent variable of a previous period.

Regression of the mortgage series with personal income lagged one quarter gives the equation

$$Y_t = \underset{(5.783)}{-30.136} + \underset{(.007)}{0.239}PI_{t-1} \qquad R^2 = .967$$

TABLE 19

LAGGED INCOME

| | $Y_p$ | $Y_p-Y_a$ | $P_i$ | $P_i-A_i$ |
|---|---|---|---|---|
| 1975:1 | 254.5 | 2.1 | 2.08 | 1.65 |
| 2 | 255.4 | -5.9 | 1.19 | -1.63 |
| 3 | 267.2 | -3.4 | 2.26 | -1.30 |
| 4 | 275.1 | -3.6 | 1.66 | -1.77 |
| 1976:1 | 282.5 | -4.1 | 1.36 | -2.29 |
| 2 | 290.6 | -9.0 | 1.40 | -4.40 |
| 3 | 297.4 | -14.7 | -0.73 | -6.44 |
| 4 | 302.5 | -20.7 | -3.08 | -8.37 |

RMS = 10.04 $\quad U^2 = 0.952$

A regression on assets lagged one quarter gives

$$Y_t = \underset{(.784)}{0.348} + \underset{(.004)}{0.868}A_{t-1} \qquad R^2 = .9991$$

Forecasts from this equation are shown in Table 20.

TABLE 20

LAGGED ASSETS

| | $Y_p$ | $Y_p - Y_a$ | $P_i$ | $P_i - A_i$ |
|---|---|---|---|---|
| 1975:1 | 256.8 | 4.4 | 3.01 | 1.77 |
| 2 | 265.0 | 3.7 | 4.99 | 1.46 |
| 3 | 276.4 | 5.8 | 5.78 | 2.22 |
| 4 | 286.0 | 7.3 | 5.69 | 2.70 |
| 1976:1 | 294.1 | 7.5 | 5.53 | 2.70 |
| 2 | 307.4 | 7.8 | 7.26 | 2.72 |
| 3 | 318.6 | 6.5 | 6.34 | 2.17 |
| 4 | 330.1 | 6.9 | 5.77 | 2.21 |
| | RMS = 6.39 | | $U^2$ = 0.443 | |

Here again, forecasts based on the asset series show better accuracy than those based on income. Results show some improvement over simple regression on the asset series, and the method is easier to use because of the elimination of the need to forecast values for the independent variable.

## E. Statistical Problems

In testing the above regressions for autocorrelation, Durbin-Watson table values for k'=1 and n=40 are $d_l$=1.44 and $d_u$=1.54. If the computed d-statistic is lower than the relevant lower table value ($d_l$), autocorrelation is present. If it is higher than the upper value ($d_u$), there is no autocorrelation. If the computed value falls between the two table values, the presence of autocorrelation is indeterminate. Computed Durbin-Watson statistics for the four regressions discussed are given in Table 21.

TABLE 21

DURBIN-WATSON TESTS

| Variable | d-value |
|---|---|
| Income | 0.095 |
| Assets | 0.916 |
| Lagged Income | 0.126 |
| Lagged Assets | 0.461 |

These results definitely show the presence of autocorrelation. To see if correcting for autocorrelation would result in more accurate forecasts, the first-differnce method was applied to these series. This involved creating a new series for each variable by substituting for each value the amount of change, or difference, from the previous value of the variable. Regressions were then run on these new series.

The regression on the first-differenced series of mortgages and income gives the equation

$$dY = \underset{(.739)}{1.284} + \underset{(.038)}{0.142}\ dPI \qquad R^2 = .268$$

TABLE 22

DIFFERENCED INCOME

| | $Y_p$ | $Y_p - Y_a$ | $P_i$ | $P_i - A_i$ |
|---|---|---|---|---|
| 1975:1 | 251.1 | -1.3 | 0.72 | -0.52 |
| 2 | 260.6 | -0.6 | 3.29 | -0.24 |
| 3 | 267.3 | -3.3 | 2.30 | -1.26 |
| 4 | 273.0 | -5.7 | 0.89 | -2.10 |
| 1976:1 | 284.8 | -1.8 | 2.19 | -0.64 |
| 2 | 291.9 | -7.7 | 1.85 | -2.69 |
| 3 | 303.9 | -8.2 | 1.44 | -2.73 |
| 4 | 302.3 | -2.9 | 2.63 | -0.93 |

RMS = 4.79 $\qquad U^2 = 0.236$

And first-difference results based on the asset series are given below.

$$dY = \underset{(.034)}{0.378} + \underset{(.064)}{0.754}\, dA \qquad R^2 = .7869$$

TABLE 23

DIFFERENCED ASSETS

| | $Y_p$ | $Y_p - Y_a$ | $P_i$ | $P_i - A_i$ |
|---|---|---|---|---|
| 1975:1 | 256.8 | 4.4 | 3.01 | 1.77 |
| 2 | 262.6 | 1.3 | 4.04 | 0.51 |
| 3 | 270.0 | -0.6 | 3.33 | -0.23 |
| 4 | 278.0 | -0.7 | 2.73 | -0.26 |
| 1976:1 | 290.6 | 4.0 | 4.27 | 1.44 |
| 2 | 296.7 | -2.9 | 3.52 | -1.02 |
| 3 | 310.0 | -2.1 | 3.47 | -0.70 |
| 4 | 320.0 | -2.4 | 2.79 | -0.77 |

RMS = 2.66 $\qquad U^2 = 0.082$

As can be seen in Table 24, the use of the first-differenced series resulted in lower error measures and better results in the Durbin-Watson test.

TABLE 24

DIFFERENCED RESULTS

| Vari. | RMS | $U^2$ | D-W | $R^2$ |
|---|---|---|---|---|
| PI | 9.61 | .872 | .095 | .97 |
| dPI | 4.79 | .236 | .859 | .268 |
| A | 7.97 | .702 | .916 | .999 |
| dA | 2.66 | .082 | 2.003 | .787 |

It should also be noted that, while accuracy tests showed improvement, the use of first differences resulted in poorer fit as measured by $R^2$. Commenting on his experience in this area, Ferber says

> ...no relationship at all was found between the coefficient of correlation of a function for the period of observation and the accuracy of the function's predictions.[1]

Therefore, the goodness-of-fit of a regression estimation does not indicate the ability of the equation to forecast future values for a series. Various independent variables must be looked at, equations estimated, and statistical corrections made. Results of these various methods are then evaluated, with the most accurate being chosen for making future forecasts. There is no a priori way of determining the best method for forecasting a series; empirical work with the series shows which method is best.

## F. Multiple Regression Models

In construction of simple models, both in economic theory and in empirical work, changes in one (dependent) variable are assumed to be correlated with changes in another (independent) variable, all other things being equal. Because other things are not equal, because other economic variables also change over time, it is difficult to isolate the effect of one variable on another.

For example, the amount of a commodity demanded is generally held to be inversely related to the price of the commodity, certeris paribus. But this relationship is difficult to estimate empirically because the demand curve shifts over time due to changes in income, tastes, etc.

---

[1] Robert Ferber, "Sales Forecasting by Correlation Techniques", Journal of Marketing XVIII (January 1954):228-229.

On this problem, Dean says

> Measurement of the pure relationship between price and sales volume, for either an industry or a company, has been a hard nut to crack statistically. The heart of the difficulty is that the setting for analysis, i.e., other determinants of demand, changes too rapidly to produce an adequately homogeneous set of data.[1]

Since other things are not constant, an alternative is to attempt to include their effect in the model. Klein says that "...it will be necessary in empirical...work to proceed from the outset with a multivariate relationship."[2] The multiple regression model allows the measurement of the correlation between a dependent variable and two or more independent variables.

Looking at the mortgage series from the demand side, the demand for mortgage funds might be taken to be a function of personal income and the price (interest rate) of the funds. Estimating this function via multiple regression gives the equation

$$\underset{(6.93)}{Y = 20.252} - \underset{(1.65)}{13.432\, r} + \underset{(.008)}{0.294\, PI} \qquad R^2 = .9897$$

where 'r' is the rate on conventional first mortgages on homes and PI is personal income. Forecasts using this model are given in Table 25.

From the supply side of the market, supply could be stated as a function of price (interest) and availability of funds (assets). A regression on these variables gives the equation:

---

[1]Dean, _op. cit.,_ p.178.

[2]Klein, _op. cit.,_ p.19.

$$Y = \underset{(1.537)}{-2.711} + \underset{(.301)}{1.075}\, r + \underset{(.005)}{0.821}\, A \qquad R^2 = .9994$$

Forecasts based on this model are given in Table 26.

TABLE 25

INCOME AND INTEREST

| | $Y_p$ | $Y_p - Y_a$ | $P_i$ | $P_i - A_i$ |
|---|---|---|---|---|
| 1975:1 | 249.5 | -2.9 | 0.08 | -1.16 |
| 2 | 265.7 | 4.4 | 5.27 | 1.74 |
| 3 | 278.8 | 8.2 | 6.70 | 3.14 |
| 4 | 287.5 | 8.8 | 6.25 | 3.26 |
| 1976:1 | 298.3 | 11.7 | 7.03 | 4.20 |
| 2 | 306.4 | 6.8 | 6.91 | 2.37 |
| 3 | 310.5 | -1.6 | 3.64 | -0.57 |
| 4 | 325.1 | 1.9 | 4.16 | 0.60 |
| | RMS = 6.72 | | $U^2$ = 0.517 | |

TABLE 26

ASSETS AND INTEREST

| | $Y_p$ | $Y_p - Y_a$ | $P_i$ | $P_i - A_i$ |
|---|---|---|---|---|
| 1975:1 | 257.4 | 5.0 | 3.25 | 2.01 |
| 2 | 268.0 | 6.7 | 6.18 | 2.65 |
| 3 | 276.8 | 6.2 | 5.93 | 2.37 |
| 4 | 284.5 | 5.8 | 5.14 | 2.15 |
| 1976:1 | 297.0 | 10.4 | 6.57 | 3.74 |
| 2 | 307.6 | 8.0 | 7.33 | 2.79 |
| 3 | 318.7 | 6.6 | 6.38 | 2.21 |
| 4 | 327.8 | 4.6 | 5.03 | 1.47 |
| | RMS = 6.88 | | $U^2$ = 0.532 | |

As with the simple regression models, these estimates show signs of autocorrelation. At 5% significance, table values for the Durbin-Watson test for k'=2 and n=40 are

$d_l=1.39$ and $d_u=1.60$. For the regression on interest and income, the computed d-statistic is 0.454; for the regression on interest and assets, it is 1.038. Therefore, the presence of autocorrelation is definitely indicated.

To see if corrective measures would improve accuracy, multiple regressions were run on the first difference series of these variables. For the regression of the change in mortgage debt on the change in interest rate and the change in personal income, the estimated equation is

$$dY = \underset{(.734)}{1.352} - \underset{(1.88)}{2.493}\,dr + \underset{(.038)}{0.151}\,dPI \qquad R^2=.301$$

For differenced interest and assets,

$$dY = \underset{(.349)}{-0.251} + \underset{(.953)}{3.267}\,dr + \underset{(.06)}{0.827}\,dA \qquad R^2=.838$$

Forecasts based on these equations are given in Table 27 and 28 below.

TABLE 27

INCOME AND INTEREST: DIFFERENCED

| | $Y_p$ | $Y_p-Y_a$ | $P_i$ | $P_i-A_i$ |
|---|---|---|---|---|
| 1975:1 | 251.9 | -0.5 | 1.04 | -0.20 |
| 2 | 261.5 | 0.2 | 3.60 | -0.07 |
| 3 | 268.3 | -2.3 | 2.68 | -0.88 |
| 4 | 276.5 | 2.2 | 2.18 | -0.81 |
| 1976:1 | 285.0 | -1.6 | 2.26 | -0.57 |
| 2 | 292.2 | -7.4 | 1.95 | -2.59 |
| 3 | 303.7 | -8.4 | 1.37 | -2.80 |
| 4 | 320.9 | -2.3 | 2.82 | -0.74 |

RMS = 4.24 $\qquad U^2 = 0.179$

TABLE 28

ASSETS AND INTEREST: DIFFERENCED

| | $Y_p$ | $Y_p-Y_a$ | $P_i$ | $P_i-A_i$ |
|---|---|---|---|---|
| 1975:1 | 255.9 | 3.5 | 2.65 | 1.41 |
| 2 | 262.6 | 1.3 | 4.04 | 0.51 |
| 3 | 269.4 | -1.2 | 3.10 | -0.46 |
| 4 | 278.7 | -0.5 | 2.81 | -0.18 |
| 1976:1 | 291.3 | 4.7 | 4.52 | 1.69 |
| 2 | 297.1 | -2.5 | 3.66 | -0.88 |
| 3 | 310.9 | -1.2 | 3.77 | -0.40 |
| 4 | 321.0 | -2.2 | 2.85 | -0.71 |

RMS = 2.51 $U^2$ = 0.072

As with the simple regression models, use of first differences rather than the original data improves the accuracy of the forecasts. With interest and income as independent variables, the RMS was lowered from 6.72 to 4.27 and the $U^2$ from 0.517 to 0.179. Using interest and assets the RMS went from 6.88 to 2.51 and the $U^2$ from 0.532 to 0.072 with the first difference method. The first difference method on interest and income gave a Durbin-Watson statistic of 1.039. This is a distinct improvement over the 0.454 of the original observations, but indicates that some autocorrelation is still present. For the first difference series on interest and assets, the computed d-statistic was 2.064, indicating virtually no autocorrelation.

Here again, the $R^2$ statistics decreased while the accuracy measurements improved, indicating that the $R^2$ is not necessarily a good indicator of the accuracy of the forecasts given by a particular estimation. Once again, the most important criteria for the selection of a forecasting technique is its past accuracy in forecasting the particular series it will be used for. If practical, a number of techniques might be constantly monitored and updated, insuring that the most accurate method is being used.

Comparing the multiple regression model results with those of the simple regression techniques, in each case the addition of a second independent variable improved the accuracy of the forecasts. These measurements are summarized in Table 29. As can be seen, the use of first differences in each case improved accuracy measures over those of the original observations.

TABLE 29

REGRESSION MODELS

| Variables | Data | RMS | $U^2$ | $R^2$ |
|---|---|---|---|---|
| Income | Orig. | 9.61 | .872 | .9700 |
| Income & Interest | Orig. | 6.72 | .517 | .9897 |
| Income | Diff. | 4.79 | .236 | .2683 |
| Income & Interest | Diff. | 4.24 | .179 | .3014 |
| Assets | Orig. | 7.97 | .702 | .9992 |
| Assets & Interest | Orig. | 6.88 | .532 | .9994 |
| Assets | Diff. | 2.66 | .082 | .7869 |
| Assets & Interest | Diff. | 2.51 | .072 | .8377 |

As a general rule in constructing regression models, the addition of another variable will improve the goodness-of-fit of the estimation, as measured by $R^2$. However, as was seen, the goodness-of-fit is not necessarily a good indicator of the accuracy of the forecasts made with a regression model. Whether the inclusion of added independent variables will improve the accuracy of the forecast, and whether the improved accuracy (if any) is worth the cost of the more complex model, must be determined in each case by empirical testing.

These methods of course by no means exhaust the possible techniques which may be used in forecasting. Other manipulations of the data may be used, such as looking at the rate of change between periods. More complex estimating techniques may be used,such as using three or more independent variables in a regression or using a simultaneous equation approach. These depend on the accuracy attained with simple methods, the degree of accuracy desired, and the constraints of time and expense under which the forecasts are made.

## CHAPTER VI

## CONCLUSION

### A. Summary of Accuracy Tests

Summary tables of accuracy test scores are given below. For testing of all naive methods, original observations of the mortgage series were used. Values used for regression models are specified for each method.

TABLE 30

NAIVE METHODS

| Method | RMS | $U^2$ |
|---|---|---|
| Continuity | 9.69 | 1.000 |
| Moving Ave. | 21.44 | 4.758 |
| E.W.M.A. | 10.63 | 1.197 |
| Abs. Change (1 period) | 2.86 | 0.098 |
| Rate of Change (1 period) | 2.89 | 0.101 |
| Abs. Change (n-periods) | 3.85 | 0.170 |
| Rate of Change (n-periods) | 2.59 | 0.083 |
| Seas. Continuity | 31.39 | 10.696 |

Neither continuity nor moving average techniques give good results in forecasting this series. These methods are more applicable to those series which do not show long term movement in one direction. If a naive method is to be used, moving average techniques are best suited to a series with fluctuation about a fairly stable mean which does not vary over time. Simple projections of change, either absolute change or rate of change, give very good results with the mortgage series. This is because of the steady long-run upward trend of this series, and these techniques would be expected to give good results with other series of this nature. Again, the major weakness of these methods in their inability to forecast turning points.

TABLE 30

REGRESSION METHODS

| Model | Data | RMS | $U^2$ |
|---|---|---|---|
| Time-Series | Orig. | 8.63 | 0.761 |
| Updated T-S | Orig. | 6.69 | 0.477 |
| Curvilinear T-S | Orig. | 7.72 | 0.674 |
| Autoregression | Orig. | 2.95 | 0.096 |
| Log Autoregress. | Orig. | 2.85 | 0.091 |
| Income | Orig. | 9.61 | 0.872 |
| Assets | Orig. | 7.97 | 0.702 |
| Income | Lag. | 10.04 | 0.952 |
| Assets | Lag. | 6.39 | 0.443 |
| Income | Diff. | 4.79 | 0.236 |
| Assets | Diff. | 2.66 | 0.082 |
| Income & Interest | Orig. | 6.72 | 0.517 |
| Assets & Interest | Orig. | 6.88 | 0.532 |
| Income & Interest | Diff. | 4.24 | 0.179 |
| Assets & Interest | Diff. | 2.51 | 0.072 |

Of the naive-type regression models, which do not use correlation with other economic variables to forecast, the autoregression methods give better results than do the time-series models for this series. For regression models using other economic series as independent variables, multiple models give better accuracy results than do the simple models, and the use of the differenced series improves accuracy over results achieved using the original data. Also, for the most part models including the asset series as an independent variable give better results than those using personal income. Although these regression are by no means exhaustive enough to form the basis for a firm conclusion, this points the the possibility that the supply side of the mortgage market is more important in determining the level of mortgage credit. It is possible that the market is determined by the availability of funds, by the proportion of total assets lending institutions hold in the form of mortgage debt, with the interest rate serving as a rationing device. A much more complex study would be needed to determine the true nature of this market.

## B. Selection of a Forecasting Technique

Obviously, there is no "best" forecasting technique. The use of one or more techniques for forecasting must be based on the nature of the particular series being forecast. Inspection of the raw data may reveal long-run growth patterns, indicating the use of trend projection methods. Frequent turning points would point to the use of leading indicators or one or more indexes to predict these turning points. Knowledge of the series and of related variables point to other economic series which may be used as independent variables in regression analysis.

The final choice of a forecasting technique must be based on empirical work with the series and with different techniques, to find which gives the most accurate forecasts. This also involves a constant update and checking process, to insure that the underlying structure of the series is not changing over time, that some technique rejected in the past may not now give better results.

# SELECTED BIBLIOGRAPHY


Balsley, Howard L. Quantitative Research Methods for Business and Economics. New York: Random House, 1970.

Bassie, V. Lewis. Economic Forecasting. New York: McGraw-Hill, 1958.

Chisholm, R. K., and Whitaker, G. R. Forecasting Methods. Homewood: Irwin, 1971.

Dauten, Carl A., and Valentine, Lloyd M. Business Cycles and Forecasting. Cincinnati: South-Western, 1974.

Dean, Joel. Managerial Economics. Engelwood Cliffs: Prentice Hall, 1951.

Dutta, M. Econometric Methods. Cincinnati: South-Western, 1975.

Elliott, J. Walter. Economic Analysis for Management Decisions. Homewood: Irwin, 1973.

Gramlich, E. M., and Jaffee, D. M. Saving Deposits, Mortgages and Housing. Lexington: Heath and Co., 1972.

Klein, Lawrence. An Introduction to Econometrics. Engelwood Cliffs: Prentice Hall, 1962.

Pindyck, Robert S., and Rubinfeld, Daniel L. Econometric Models and Economic Forecasts. New York: McGraw-Hill, 1976.

Spencer, Milton H., and Siegelman, Louis. Managerial Economics. Homewood: Irwin, 1959.

Theil, Henri. Applied Economic Forecasting. Amsterdam: North-Holland, 1966.