SPAINHOUR, LARRY SAMUEL. An Application of Forecasting Techniques for a Small Geographical Area. (1977) Directed by: Dr. G. Donald Jud. Pp. 79

It was the purpose of this thesis to formulate and test short-run economic forecasting methodologies that are useful for small geographical areas. The major concern was to derive an accurate monthly revenue forecast for the Chesapeake and Potomac Telephone Company of Washington, D.C.

A survey was made of the literature dealing with techniques used to forecast telephone demand. This review suggested methodologies that were appropriate given the special problems of the Washington area. A narrative analysis of current economic trends affecting telephone demand in Washington further refined the development of a proper forecasting methodology.

Next, an analysis of the available time series data was presented. This analysis provided an understanding of the underlying characteristics of the data and led to the formulation of specific forecasting models for testing.

Five separate empirical models were developed for forecasting telephone demand in Washington, D.C. These models were analyzed for their statistical significance and tested for their ability to produce accurate forecasts. Monthly forecasts were generated with each model, and a final forecast was selected.

AN APPLICATION OF FORECASTING TECHNIQUES

FOR A SMALL GEOGRAPHICAL AREA

by

Larry Samuel Spainhour

A Thesis Submitted to
the Faculty of the Graduate School at
The University of North Carolina at Greensboro
in Partial Fulfillment
of the Requirements for the Degree
Master of Arts

Greensboro
1977

Approved by

G. Donald Jud

Thesis Adviser

APPROVAL PAGE

This thesis has been approved by the following commitee of the Faculty of the Graduate School at the University of North Carolina at Greensboro.

Thesis Adviser G. Donald Jud

Committee Members Gary T. Barnes

James M. Watts

May 2, 1977

Date of Acceptance by Committee

ACKNOWLEDGMENTS

I would like to express my appreciation to Dr. G. Donald Jud, Department of Economics, for his guidance in the research and formulation of this thesis. His many helpful comments and suggestions, along with those of Drs. Gary T. Barnes and James M. Watts, were invaluable.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Econometric models are used by both the government and private industry to explain and forecast aggregate economic conditions. Economists develop econometric models in order to explain economic concepts like the consumption function and various economic activity indices. Econometric models are also used at the micro level of economics by private industry to forecast things like product demand. Such models, when used to forecast, are usually of an aggregate nature, that is, they apply to an industry as a whole or to some large firm.

Many forecasts are needed in private industry at a level much smaller than the entire industry or firm. These needs create different problems for the econometrician who is interested in providing accurate forecasts and not, necessarily, in explaining economic phenomenon.

It is this problem of making accurate forecasts for a small geographical area that is the topic of this thesis. In dealing with the problem, the ultimate objective is to provide a revenue forecast for the Chesapeake and Potomac

Telephone Company (C&P) of Washington, D.C., a subsidiary of American Telephone and Telegraph.

The Bell System is a public utility, and as a public utility, the system has a commitment to the public to provide access to and the use of the telecommunications network upon customer demand. In this respect, Bell System services are not governed by supply and demand in the way that most goods and services are. Because of this public commitment, the supply of Bell System services is driven by customer demand, that is, the Bell System is obligated to supply all services demanded at a price which is sufficient to allow the company to earn a reasonable rate of return on its investment.

Customer demand is ever changing in respect to time and location, and it is impractical to place facilities in all possible locations where customers are likely to locate. Therefore, in order to meet its commitment to the public and in order to operate at maximum efficiency, the Bell System attempts to forecast the demand for its services and in turn the revenues that will be generated by that demand.

The Bell System includes twenty-one individual operating companies of which the Chesapeake and Potomac Telephone Company of Washington, D.C., is one. Each of these companies is responsible for supplying its area with the service demanded. Each is also required to forecast the demand for services and the revenues that will be generated by this demand.

As previously stated, the ultimate goal of this thesis is to forecast revenue for the C&P of Washington, D.C. As the name of the company implies, it is responsible for providing services to the city of Washington, D.C.

The forecasting models presented here are concerned with the monthly charges portions of local service revenues. Monthly charges revenues are payments received on a recurring basis for the customers' exchange service. These revenues are based on the amount and type of equipment that the customer selects and has installed on his premises. The revenues received depend on a number of things, such as the type of telephone, the number of lines and extensions, the class of service (resident or business), and the type and number of listings. Hereafter, monthly charges revenues are referred to as MCR.

Before describing the methodology selected to forecast MCR for the city of Washington, it is appropriate to review some of the works of others and discuss some of the techniques that have been used in forecasting telephone demand. This is the subject of chapter 2. Chapter 3 consists of an analysis of MCR historical data. The analysis identifies the underlying characteristics of the data and provides some direction for model development. Chapter 4 formulates and tests some statistical models for forecasting telephone demand in Washington, D.C. The models are analyzed for significance and tested for forecasting accuracy, and forecasts are generated with each model. Finally, in chapter 5 the work presented is summarized and a final forecast is selected. The thesis is concluded with a few remarks concerning the direction of future investigations.

CHAPTER II

ANALYSIS OF TELEPHONE DEMAND: A REVIEW

Literature Review

As was mentioned earlier, the demand for telephone service determines the amount of service to be supplied; therefore, it is reasonable to assume that revenues received for service are a function of the demand for service. That is, supply in the short run is perfectly elastic at the going price. For this reason it is important to emphasize the demand for service in this study. There is not a great deal of literature devoted to the demand for telephone service. Compared to other industries, such as the automobile industry and even to other public utilities like electrical energy, the literature concerned with the telephone industry is extremely scarce. No attempt is made here to explain why this situation exists, although it does seem strange in light of the fact that the telephone industry is one of the largest industries in this country.

The first work to be mentioned here was written by B. E. Davis, G. J. Caccappolo and M. A. Chaudry. These men developed an econometric planning model for American Telephone and Telegraph Company. In their model, they incorporated both a "demand module" and local service revenue module.

In the first place, it is important to make note of their strong emphasis on demand and external economic conditions. This emphasis is evidenced by their statements that,

> The overall modeling approach is based on the premise that the state of the economy determines an individual firm's demand, making it externally derived rather than created by the firm's supply capability. With this mode dominating, Bell System demand is assumed dependent upon economic factors external to the Corporation and supply is a reaction, via corporate policy actions, to the demand.[1]

Their model is an aggregate model for the entire Bell System, and the driving force is a forecast of the state of the country's economy. As stated, their model includes a demand module, and in the demand module they include four exogenous variables. These variables consist of some measure of the state of the national economy, demography, prices, and consumer tastes. In the section devoted to revenues, they assume local service revenue to be a function of a local service price index, the implicit GNP deflator, personal disposable income per capita less government transfers to persons, an implicit deflator for personal consumption, and total telephones excluding residential extensions.

If one could assume that the economy of the city of Washington is similar in nature to the economy of the country, one could develop a model using these variables for forecasting in the District. However, it is unreasonable to make such an assumption. This idea will be discussed in more detail a little later in this chapter.

In other work, it has proven successful to divide the aggregate and model the subaggregate. This idea seems very practical. For example, it stands to reason that the determinants of the demand for residence telephones are not the same as those for business telephones; therefore, one can probably better forecast total demand by modeling both residence and business demand. Such an approach is taken by Roshan L. Chaddha and Sharad S. Chitgopekar.[2] They use three variables to model residence telephones. In their model they assume that the demand for residence telephones is a function of the number of households, per capita disposable income, and revenue per telephone.

In the Bell System this idea of forecasting subaggregate items has been effectively employed in forecasting stations, that is, residence and business telephones. However, it is impossible under present circumstances to employ this method in revenue forecasting because revenues are not currently collected and reported in such a subaggregated nature.

Another work worthy of note is written by Douglas M. Dunn, William H. Williams and Allen W. Spivey.[3] They not only employ the use of a subaggregate method of forecasting, but they also make some interesting comparisons between models using exogenous variables and models that do not. In their analysis, they are interested in modeling residence telephone demand for Flint, Michigan. They develop a model using a monthly count of the total number of employees covered by unemployment compensation and a yearly series of the total number of households. They compare this model with some prior models that have been used for Flint which did not use exogenous variables. Their results show that the model using the two exogenous variables produces a 15 percent gain in forecast accuracy over the best of the prior models.

These are very significant results, but equally important are some other comments concerning the use of exogenous variables in small geographical areas. In their analysis, Dunn, Lewis and Spivey note,

> In forecasting at the level of the national economy or of the total Bell System, the geographical context may not be an important factor. As the geographical area becomes smaller, however, one must take into account this influence. Exogenous data become much harder to obtain, and short-term local swings caused by strikes, welfare policy, and changing deposit practices appear to be impossible to predict by analytic techniques.[4]

These comments are very relevant to the problem of forecasting MRC. Not only must adequate historical data exist for the proper exogenous variables, but in most cases one must have a reliable source of forecasts of these variables. This is necessary unless the variables lead the dependent variable by a considerable length of time. By experience it can be said that the latter is seldom the case.

Dunn et al. also make mention of two other major problems in relation to the use of exogenous variables. First, "one has to decide which of all the conceptually useful exogenous variables are likely to be fruitful," and second, "these data must be obtained with a reasonable expenditure of time and resources."[5]

As an alternative to the use of exogenous variables, or at least to supplement such models, Dunn et al. recomment the use of exponential smoothing, autoregression and spectrum analysis. The first two of these three will be discussed in detail in chapter 4.

The literature suggests that, given the proper circumstances, models which include economic variables are superior to those which do not. However, for Washington, the circumstances are not such that this premise holds true. Attempts have been made by previous members of the C&P forecasting staff and by the business research staff to develop econometric models for forecasting revenues for the District. All attempts have proven to be unsuccessful. There are two basic reasons for these failures.

First, there is the problem of geography. As was mentioned earlier, when forecasting a small geographical area, what would be a minor influence on a larger area can have a tremendous effect on a small area. This is best explained by example. In 1968, the city of Washington experienced a violent civil disturbance. The disturbance had a tremendous effect on business activity. Needless to say, it is impossible to predict an influence of this nature. It is not an easy task to quantify the effects of such influences even after they occur. Another example might involve the move of a governmental agency or large business concern out of the area. Influences such as these could have very little or no effect on the national economy or some other large geographical area; however, they are enough to severely damage the accuracy of a forecast for a small area like Washington, D.C.

In small geographical areas reliable economic data are also difficult to obtain. This is true of the city of Washington. In many cases, adequate economic data exist for the Washington metropolitan area but not for the District itself. Where adequate historical data do exist, one cannot obtain reliable forecasts of the data. This is true, for example, of unemployment data. The historical data are available but no one is able to provide adequate forecasts of future unemployment rates.

The second reason for the lack of success of econometric models involves the unique nature of the city's economy. It is reasonable to use national economic data in making forecasts for smaller areas where the area's economy has many of the same characteristics as does the national economy. If a small area has a cross section of the nation's industry, then trends in economic activity in the small area might parallel those of the nation. If such is the case, it is reasonable to use national economic data to forecast for that particular area.

One does not have to look far to realize that the economy of Washington is nothing like that of the nation. More broadly, it can be said that the nation's capital is unlike any other place in the country. As anyone could guess, its economy is dominated by government. This fact more than any other makes the city a difficult place to monitor and, in turn, a difficult place to forecast. Our government has the power to react swiftly, when required, and agencies might undergo rapid expansion. Yet our government is big and awkward, and what were firm plans yesterday can be postponed for years.

On the other hand, the dominance of government acts as a stabilizer for the District's economy. It does not experience large swings of recession and expansion as does the rest of the nation's economy. In good times and bad, the federal government payrolls change very little. The last few years demonstrate this stabilizing effect. While the nation suffered through the worst recession in over forty years, the District's economy only leveled off and dipped slightly.

For the above reasons, econometric models with economic variables are not developed here. It is considered that they would not only be unsuccessful but a waste of money and energy. Instead of trying to develop econometric models, the remainder of this chapter examines specific variables that affect telephone demand.

Current Trends in Washington, D.C.

The residence side of the market is examined first. Obviously, it is related to population and the number of households. The population of the District experienced rapid growth until the late 1960s, reaching a high of approximately 794,000 in 1966. However, with the racial disturbances of the late 1960s, the population started to decline. Recent figures released by the Bureau of Census placed the District's population at 702,000 as of July 1, 1976. This represents a decline of more than 11 percent. According to C&P's Business Research organization, this decline is expected to continue for the short term.

While the population has been declining, households have been increasing. It is amazing that households have been able to maintain any growth, considering the decreases in population. This has occurred because of the rapid decline in the size of the family unit. Since it is households that demand residence telephones, one might expect continuing slow growth in the number of residence telephones during the next few years.

These trends in population and households result from a lack of residential construction activity. Residential construction has been down sharply since the racial disturbances. Burned apartment buildings have been boarded up and are not being replaced. Old buildings are being shut down instead of being renovated. Developers are not willing to accept the risks associated with building within the District. Lending organizations are directing their money into the suburbs of Maryland and Virginia, where the demand for housing is high. All of these factors exist because of deteriorating social conditions inside the city.

The residence market accounts for only about 25 percent of local service revenue, with the other 75 percent coming from business. While the business market is substantial, it lacks growth potential. Many businesses have followed the movement of population toward the suburbs. This is particularly true for those businesses involved in service industry. The lack of adequate office space, transportation difficulties, and expensive parking make the metropolitan area outside the city more attractive to business. In the long run, the completion of the subway system could reverse this trend; however, the system is running into mounting financial problems, and there are serious questions as to whether the system will ever reach the fringe areas of Northern Virginia.

There are a few large programs being started that are designed to revitalize the city. One such plan is the "Pennsylvania Avenue Plan," which will restore the thoroughfare between the Capitol and the White House. The plan calls for construction of both residential and commercial facilities. The plan was authorized by the Congress in August 1976; however, it was first proposed more than fifteen years ago and is definitely a long-range project.

Another large project involves the reconstruction of the Fourteenth Street corridor, which was burned out during the racial disturbances. This project has also been slow getting off the ground and no immediate growth will be realized.

These discussions do not tell one what revenues will be in 1977, but they help establish the environment of the area in which C&P operates. They suggest that one cannot expect any extraordinary growth in 1977. No new markets will be available, and existing markets are rather saturated. With the above background, chapter 3 proceeds with an analysis of the historical data, while chapter 4 develops a series of statistical models suitable for forecasting.

FOOTNOTES

[1]B. E. Davis, G. J. Caccappolo, and M. A. Chaudry, "An Econometric Planning Model for American Telephone and Telegraph Company," The Bell Journal of Economics and Management Science 4 (Spring 1973):30.

[2]Roshan L. Chaddha and Sharad S. Chitgopekar, "A Generalization of the Logistic Curves and Long-range Forecasts (1966-1991) of Residence Telephones," The Bell Journal of Economics and Management Science 2 (Autumn 1971):542-60.

[3]Douglas M. Dunn, William H. Williams, and Allen W. Spivey, "Analysis and Prediction of Telephone Demand in Local Geographical Areas," The Bell Journal of Economics and Management Science 2 (Autumn 1971):561-76.

[4]Ibid., p. 562.

[5]Ibid., p. 573.

CHAPTER III

ANALYSIS OF AVAILABLE TIME SERIES DATA

Table 1 contains actual MCR which have been received by C&P of Washington. The data span the period January 1960 through July 1976. In order to use these data for forecasting purposes, it is necessary that a number of adjustments be made. Most important are the adjustments for price effects. The data must be expressed on a current price basis. As the telephone industry is a public utility, telephone rates are regulated by the state or local government. Changes in rates must be approved by a public utility commission before they are placed into effect. When the commission approves a rate increase, all historical data must be increased in order to maintain a common base. Because price is regulated and independent of variables which affect quantity demanded, revenues need to be adjusted so that they are always proportional to quantity. This is generally accomplished by determining the percentage increase realized by the rate case and applying that increase to the historical data. For example, if the rates are increased by 5 percent, then all previous data points should be multiplied by 1.05. It is not necessary to list

TABLE 1

ACTUAL MONTHLY CHARGES REVENUES
(In thousands of dollars)

Month	Year								
	1960	1961	1962	1963	1964	1965	1966	1967	1968
Jan	2,468	2,691	2,848	3,062	3,300	3,538	4,089	4,392	4,664
Feb	2,486	2,687	2,879	3,088	3,308	3,588	4,068	4,455	4,672
Mar	2,510	2,715	2,886	3,114	3,332	3,604	4,116	4,452	4,713
Apr	2,511	2,718	2,881	3,132	3,346	3,609	4,186	4,480	4,764
May	2,533	2,749	2,911	3,142	3,368	3,642	4,200	4,478	4,791
Jun	2,552	2,741	2,932	3,153	3,304	3,695	4,214	4,524	4,832
Jul	2,569	2,762	2,933	3,169	3,430	3,712	4,349	4,464	4,832
Aug	2,583	2,774	2,969	3,192	3,382	3,811	4,289	4,589	4,903
Sep	2,644	2,807	2,983	3,207	3,457	3,964	4,272	4,577	4,895
Oct	2,684	2,804	2,998	3,262	3,484	3,979	4,340	4,629	5,047
Nov	2,670	2,818	3,027	3,250	3,490	4,030	4,424	4,665	5,024
Dec	2,651	2,865	3,041	3,298	3,528	4,037	4,401	4,691	5,049

Month	Year								
	1969	1970	1971	1972	1973	1974	1975	1976	
Jan	5,043	5,281	5,511	5,964	7,154	7,147	8,049	8,308	
Feb	5,104	5,289	5,611	6,011	6,972	7,309	8,181	8,547	
Mar	5,106	5,287	5,600	6,039	7,106	7,275	8,025	8,385	
Apr	5,114	5,436	5,602	6,013	7,120	7,244	7,871	8,394	
May	5,133	5,390	5,625	6,052	6,985	7,371	8,414	8,348	
Jun	5,180	5,422	5,689	6,141	6,967	7,317	7,950	8,865	
Jul	5,188	5,384	5,786	6,148	7,051	7,335	8,030	8,958	
Aug	5,262	5,359	5,866	6,899	7,072	7,317	8,033		
Sep	5,258	5,364	5,689	6,848	7,068	7,388	8,048		
Oct	5,307	5,386	5,803	6,799	7,196	7,469	8,102		
Nov	5,327	5,567	5,902	6,936	7,083	11,445	8,182		
Dec	5,348	5,391	5,965	6,766	7,237	8,141	8,280		

all such adjustments that were required; however, it is interesting to note that for the period which the data span, adjustments were required for five rate increases and four rate decreases. The important concept of price adjustments is that, when a rate change takes place, the adjustment is made over all data points that precede the change.

It is important at this time to mention a rate change that will be handled in a somewhat unusual manner. C&P of Washington was granted a rate increase effective June 1, 1976. Rate changes are involved and complicated because they affect various types and classes of service. It is desirable to make rate adjustments only after one can analyze a number of months of data which include the new rates. For this recent rate case, a different approach is taken. Temporary adjustments are made for June and July which act to negate the rate increase. This means that the forecasts that are derived from the data base exclude the rate increase that is currently in effect. This fact is compensated for and discussed in more detail in chapter 5.

In addition to rate or price adjustments, the need arises on occasion for one-time adjustments. These adjustments generally affect only one month but at times are required over an extended period. For MCR the needs for this type adjustment are numerous. They include such things as billing errors, unusual credits and carrying charges and are generally supplied by the accounting organization.

After making the necessary adjustments to the data base, data analysis provides some insight into the underlying characteristics of the revenues to be forecast. However, before continuing, the underlying characteristics of any time series data are discussed. Generally, time series data are considered to consist of four components: trend, cycle, seasonality and irregularity. Trend is best defined as the central tendency or direction in which the data are heading. Trend is commonly discussed in terms of the slope of the data, be it positive, negative or flat. Cycle is the recurrence of a particular pattern in the data that generally occurs over a long period of time (more than one year). Cyclical patterns are usually associated with the "business cycle." Seasonality is variation or fluctuations in the data that tend to repeat themselves in patterns that follow the changing seasons of the year. These patterns are generally associated with the weather and are affected by the major holidays. Irregularity is the unexplained component of the data that follows no known reason, rule or pattern. Irregularity cannot be explained and cannot be forecast.

Having defined the four components of time series data, it is important to be able to measure the relative dominance each has in the data. This can be accomplished by performing an analysis of variance on a two-way table of the data, in which the columns are identified by years and the rows by months. Table 2 contains such a table for base adjusted MCR. Performing an analysis of variance on these data yields the information contained in table 3. With this information the hypothesis that seasonality and trend and cycle do not exist is tested. In order to reject this hypothesis, F-statistics greater than approximately 1.90 are required at the 95 percent significance level. Given the F-statistics in table 3, one can conclude that there is significant seasonality, and trend and cycle.

Knowing that these components are significant, it is now important to know how dominant they are. This is accomplished by taking ratios of the sums of squares. Dividing the sums of squares for rows (1,983,500) by total variation (520,765,680), a measure of the amount of variation explained by seasonality is obtained. This measure is computed to be .4 percent. Doing this for columns, the amount of variation explained by trend and cycle is found to be 99.5 percent. The remaining variation, .1 percent, is attributed to irregularity. One can conclude from this information that, while seasonality is a significant component, the data are dominated by trend and cycle.

The above information provides some insight into how to begin to model this revenue series. A regression on time is a reasonable starting point, given the dominance of

TABLE 2

BASE ADJUSTED MONTHLY CHARGES REVENUES
(In thousands of dollars)

Month	Year								
	1960	1961	1962	1963	1964	1965	1966	1967	1968
Jan	2,936	3,201	3,388	3,649	3,932	4,216	4,658	5,025	5,340
Feb	2,957	3,197	3,425	3,680	3,942	4,276	4,703	5,098	5,391
Mar	3,986	3,230	3,433	3,711	3,971	4,295	4,710	5,094	5,396
Apr	2,987	3,233	3,427	3,732	3,987	4,301	4,790	5,126	5,455
May	3,013	3,270	3,463	3,744	4,013	4,340	4,806	5,124	5,486
Jun	3,036	3,261	3,488	3,757	4,039	4,403	4,822	5,180	5,533
Jul	3,056	3,286	3,489	3,776	4,087	4,423	4,907	5,226	5,533
Aug	3,073	3,300	3,532	3,804	4,030	4,378	4,908	5,254	5,614
Sep	3,145	3,339	3,554	3,822	4,120	4,516	4,888	5,241	5,605
Oct	3,193	3,336	3,573	3,887	4,152	4,533	4,966	5,300	5,737
Nov	3,176	3,352	3,607	3,873	4,159	4,591	4,993	5,341	5,752
Dec	3,154	3,408	3,624	3,930	4,204	4,599	5,036	5,371	5,781

Month	Year								
	1969	1970	1971	1972	1973	1974	1975	1976	
Jan	5,774	6,138	6,506	6,992	7,491	7,608	7,905	8,342	
Feb	5,844	6,148	6,618	7,047	7,421	7,780	8,097	8,516	
Mar	5,846	6,145	6,569	7,080	7,404	7,744	7,962	8,525	
Apr	5,855	6,318	6,571	7,050	7,419	7,711	8,113	8,532	
May	5,877	6,265	6,598	7,095	7,435	7,846	8,193	8,407	
Jun	5,931	6,401	6,673	7,200	7,416	7,789	8,115	8,548	
Jul	5,940	6,356	6,787	7,208	7,505	7,808	8,110	8,761	
Aug	6,025	6,327	6,816	7,271	7,528	7,789	8,226		
Sep	6,020	6,332	6,761	7,368	7,523	7,841	8,146		
Oct	6,076	6,358	6,807	7,369	7,660	7,923	8,292		
Nov	6,099	6,390	6,923	7,369	7,593	7,946	8,248		
Dec	6,216	6,546	6,974	7,349	7,703	7,997	8,283		

trend. However, before proceeding with this, it is appropriate to examine a plot of the data. (See figure 1.) Looking at the plot, one can see the dominance of trend. Going a step further, it can be said that the trend appears to be linear, which suggests that the regression on time should be linear. Still looking at the plot of the data, an appropriate regression period must also be selected. The idea, of course, is to select the most current period which appears to best represent the data. For MCR, this period appears to be January 1965 through July 1976.

TABLE 3

ANALYSIS OF VARIANCE

	Sums of Squares	Degrees of Freedom	Mean Square	F-Statistic
Rows[a]	1,983,500	11	180,318	95.31
Columns[b]	518,470,000	15	34,564,666	18,268.85
Residual[c]	312,180	165	1,892	
Total Variation	520,765,680	191	34,747,876	

[a]Variation due to seasonality.

[b]Variation due to cycle and trend.

[c]Unexplained variation.

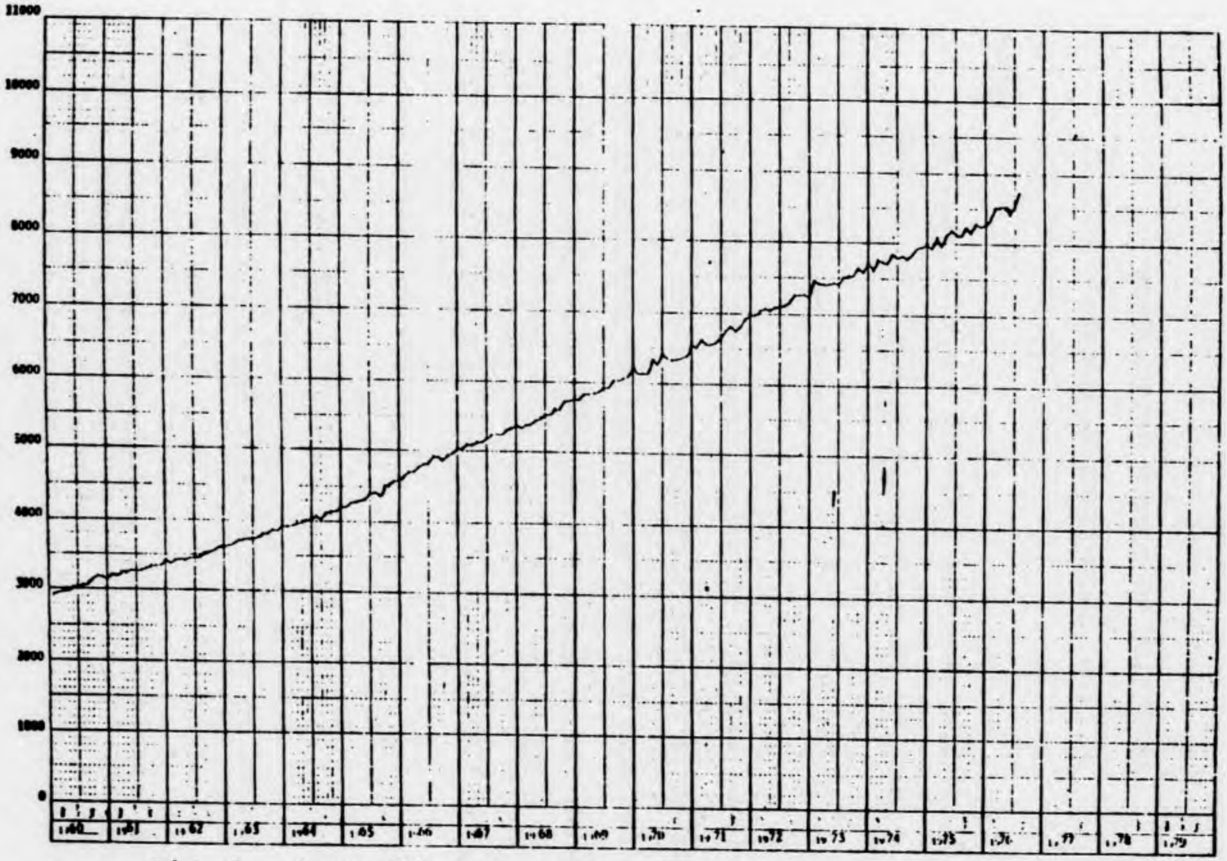

Fig. 1. Base adjusted monthly charges revenues (in thousands of dollars).

CHAPTER IV

FORECASTING MODELS OF TELEPHONE DEMAND IN WASHINGTON, D.C.

Model 1

As stated in the previous chapter, the first model to be evaluated is a linear regression of the form

$$Y_t = A + BX_t + \mu ,$$

where Y_t is the dependent variable, A is the population constant, B is the regression coefficient, X_t is some level of the independent variable, and μ is the disturbance term. The disturbance term represents the sum of all the neglected influences on the dependent variable. The fitting technique which is used to estimate this equation is the least-squares criterion where

$$Y_t = a + bX_t + e_t .$$

In this equation, a and b are statistics of estimates of A and B.

> The least-squares technique yields estimates which have the property that the sum of squares of the residuals, for a sample, are minimized. A residual is the difference between an actual and a fitted value of the

dependent variable, where the fitted value for period T is given by $(a + bX_t)$.[1]

The disturbance term is assumed to have an expected value of zero.

Least-squares equations for estimates of a and b are as follows:

$$a = \frac{\sum Y_i - b\sum X_i}{N}$$

$$b = \frac{N\sum X_i Y_i - \sum X_i \sum Y_i}{N\sum X_i^2 - (\sum X_i)^2}$$

The first model is a regression of MCR on time where the regression period is January 1965 through July 1976, and time is measured from a value of 23568 in the first period (January 1965) to a value of n in the last period. For this model n equals 23706. There is no particular reason that time should start with 23568. In this case, values for time are generated by the software.

Applying the least-squares equations for estimating a and b to the data, the following equation is derived:[2]

$$M_t = -73732 + 31.466\ T_t$$

This equation can be used to forecast future levels of MCR; however, a forecast is not in order until various tests of significance are performed on the equation.

Before looking at the tests of significance for this equation, it is important to note that of primary importance to statistical estimating and testing is the nature of the disturbance term. While the term is assumed to be zero, there are always influences that are too small to be dealt with or cannot be pinpointed.

> It is the presence of this term in a model that makes regression analysis a stochastic or probabilistic study rather than one of exact measurement. A disturbance term changes a deterministic economic model into a stochastic, econometric model. The disturbance term is assumed to have certain properties in order to carry out statistical estimation and tests of significance.[3]

The following assumptions apply to the disturbance terms:

1. The first assumption, as previously stated, is that the disturbance term is expected to have a value of zero.
2. All values of the disturbance term are assumed not to be correlated with one another.
3. All values of the disturbance term are assumed to have a constant variance.
4. The disturbance term is assumed not to be correlated with the independent variable(s) in the equation.
5. Finally, it is assumed that the disturbance term is normally distributed; that is, the frequency distribution of μ is described by the normal curve of error.[4]

The first test of significance for the estimated linear regression equation will be to examine the amount of the variation in the data that is explained by the equation. This is accomplished by computing the square of the correlation coefficient, R^2. This statistic, as it is commonly stated, measures the "goodness of fit" of the regression equation to the data. The equation for computing R^2 is as follows:

$$R^2 = \frac{\sum_{t=1}^{n}\left(\hat{Y}_t - \overline{Y}\right)^2}{\sum_{t=1}^{n}\left(Y_t - \overline{Y}\right)^2}$$

In this equation the numerator measures the variation of Y explained by the regression equation and the denominator measures the total variation of Y. The quotient measures the amount of the total variation of Y that is explained by the regression equation. For the estimated linear regression equation R^2 equals 99.7 percent, which says that 99.7 percent of the variation in observed MCR is explained by the regression equation.

Before proceeding, it should be said that the value of the R^2 statistic can be misleading to the forecaster. Because of its definition, the forecaster could mistakenly interpret the R^2 to reflect the forecasting ability of the equation. A common mistake is to compare the R^2 values of different models and make a judgment as to which model would produce the best forecast. "Actually the value of R^2 increases for a given set of sample data as more variables are added to the equation, regardless of the relevance of these variables."[5] The matter of evaluating the forecasting accuracy of a model is discussed in more detail later.

The next test consists of an evaluation of the significance of the estimated coefficients of the linear regression equation. Repeating the equation as

$$M_t = -73732 + 31.466\ T_t$$
$$(3360.9) \quad (.14219)$$

the numbers in parentheses are standard errors of the regression coefficients. By dividing the regression coefficients by their standard errors, t-statistics are obtained which are used to evaluate the significance of the estimated coefficients. The computed t-statistics for this equation are -219.38 for the constant term and 221.30 for the coefficient for time.

First, the null hypothesis is tested that

$$H_0 : a = 0 ,$$

that is, that the dependent variable is proportional to the independent variable or, in other words, the dependent variable is zero when the independent variable is zero. This hypothesis is tested against the alternative hypothesis that

$$H_1 : a \neq 0 .$$

Testing these hypotheses, one compares the computed t-statistic for the constant term, -219.38, to the 95 percent significance level of t for 137 degrees of freedom.

This tabular value is found to be ±1.98. Since the computed t-statistic is greater than the tabular value, one can reject the null hypothesis that the constant term is equal to zero.

It has not been previously stated, but by simply looking at a plot of MCR, one can assume that time should have a positive effect on M. Therefore, one can test the one-sided hypothesis:

$$H_1 : b > 0 ,$$

that is, that the regression coefficient for time is greater than zero and thus has a positive effect on M. In order to test this hypothesis, one compares the computed t-statistic of 221.30 to the tabular t-value of 1.66. Since the computed value is greater than the tabular value, one can accept with 95 percent confidence the hypothesis that the regression coefficient for time is greater than zero and is therefore significant.

Having evaluated the estimated regression coefficients and determined that both are significant, the next test is to evaluate the regression equation as a whole. This is accomplished with an F-test, where

$$F = \frac{R^2 / (K - 1)}{(1 - R^2) / (n - K)}$$

and K is the number of regression coefficients in the model.

The computed F-statistic for the linear regression equation is 48,972.539. This value is compared to a critical value ($F^{.05}_{1,137}$) of approximately 3.92. Since the computed value is greater than the critical value, one can assume that the estimated regression equation is significant with 95 percent confidence.

To this point the linear regression equation appears to be very sound as all tests have demonstrated that the model is significant; however, two important tests remain. The first involves a visual examination of a plot of the residuals. A residual is the difference between an actual value of the dependent variable and a value that is calculated by the equation. If all the assumptions of the error term are met, the residuals will appear to be random about the calculated values with a constant variance, independently related to one another and not correlated with one another.

Figure 2 is a plot of the residuals for the linear regression model. Looking at this plot, one cannot say that the residuals are random. They appear to have a cyclical pattern, which means that they are correlated to one another and are thus dependent on one another. When this is true, autocorrelation is present. This does not necessarily mean that the model should not be used for forecasting; however, it does make tests of significance using t-statistics and F-statistics questionable. The

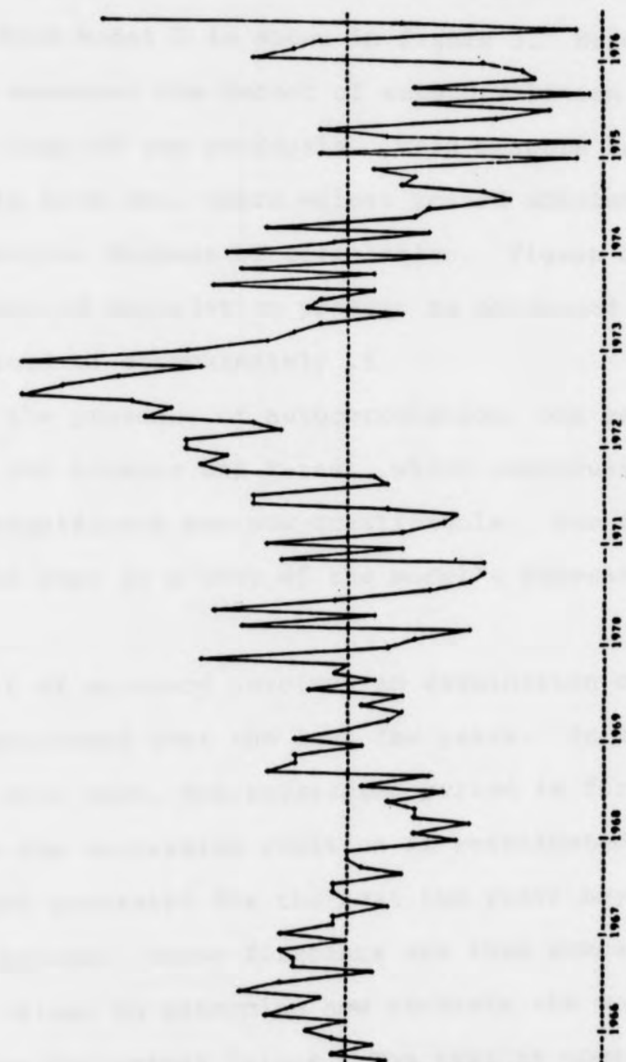

Fig. 2. Model 1--residual plot (January 1965 through July 1976).

presence of autocorrelation can be verified with an autocorrelation plot of the residuals. Such a plot for the residuals from Model 1 is shown in figure 3. Here the vertical axis measures the amount of autocorrelation present at various lags of the residuals. This measure ranges from minus one to plus one, where values toward absolute one represent higher degrees of correlation. Figure 3 reflects a high degree of correlation present as evidenced by several values of approximately .5.

With the presence of autocorrelation, one can conclude that the t-tests and F-test which previously proved the model significant are now questionable. One final test remains, and that is a test of the model's forecasting ability.

A test of accuracy involves an examination of how the model has performed over the last few years. In order to accomplish this test, the regression period is first shortened. Then the regression equation is reestimated, and forecasts are generated for the next two years beyond the regression period. These forecasts are then compared to the actual values to determine how accurate the model is in predicting the actual values. The test is most appropriately performed over a period of at least four years. First the equation is estimated using the period January 1965 through December 1971. Forecasts are generated for the period January 1972 through December 1973. These

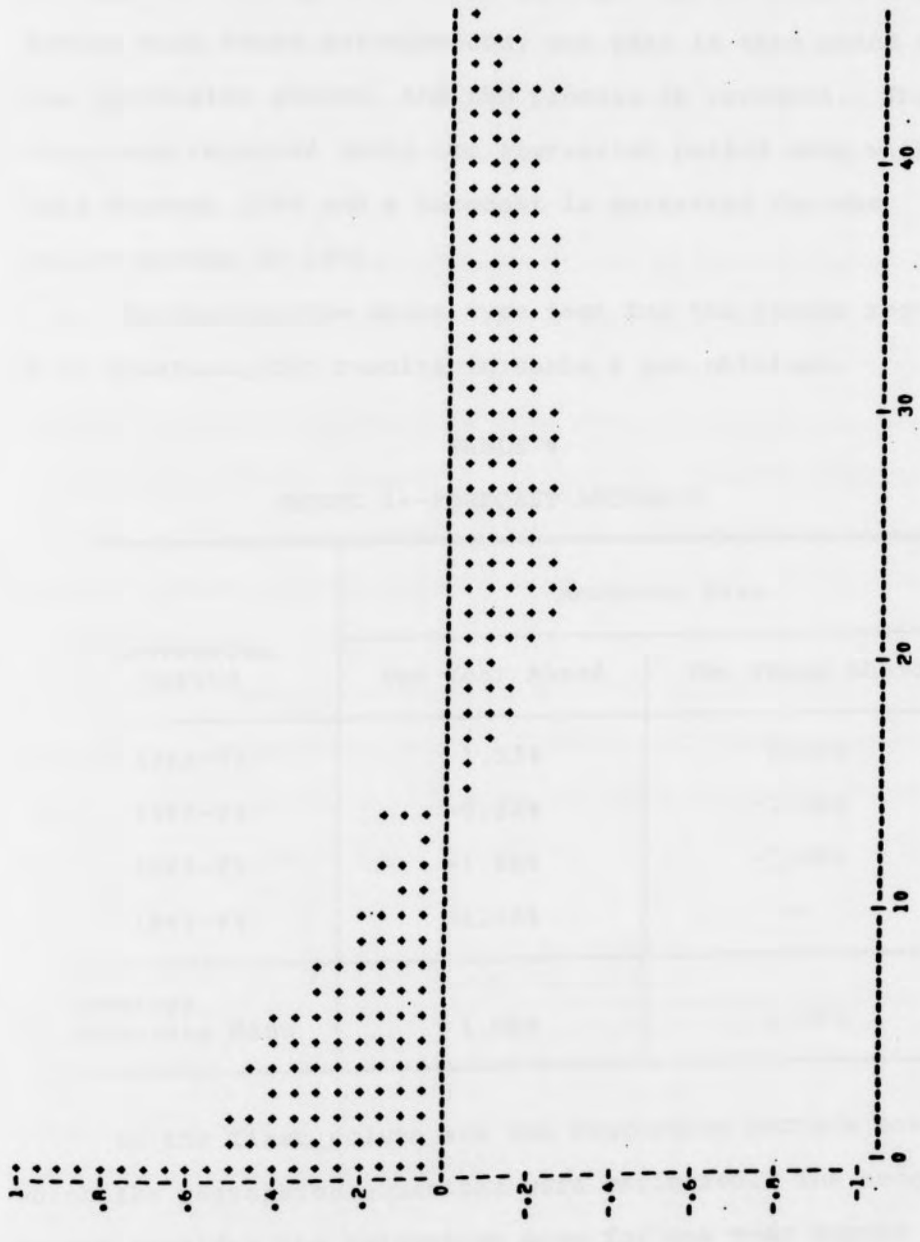

Fig. 3. Model 1--autocorrelation of residuals.

forecasts are then compared to the actuals for that period, and from this comparison a percentage miss is calculated. Having made these calculations, one year is then added to the regression period, and the process is repeated. These steps are repeated until the regression period ends with data through 1974 and a forecast is generated for the twelve months of 1975.

Performing the above type test for the linear regression equation, the results in table 4 are obtained.

TABLE 4

MODEL 1--FORECAST ACCURACY

Regression Period	Accuracy Miss	
	One Year Ahead	Two Years Ahead
1965-71	1.53%	0.52%
1965-72	-0.22%	-1.26%
1965-73	-1.16%	-1.88%
1965-74	-1.43%	--
Average Absolute Miss	1.08%	1.22%

In the first column are the regression periods over which the regression equations were estimated. The second column provides the percentage miss for one year beyond the regression period. Column three provides the same for the second year beyond the regression period. A positive miss is an under forecast and a negative miss is an over forecast. From the data in table 4, one can see that the model has over forecast the last three years, and each one by an increasing amount. In accordance with recommendations from the AT&T Forecasting Organization, an acceptable miss is .50 to 1.00 percent. This model is considered to have poor forecasting ability at 1.08 and 1.22 percent. Since the model is using data through July 1976, the forecast period of January through December 1977 is six through seventeen months beyond the regression period. The expected forecast miss is about 1.15 percent, the midpoint between 1.08 and 1.22 percent.

Having evaluated this model, the final step is to generate a prediction using the estimated regression equation. Applying the equation to calendar year 1977, the prediction in table 5 is obtained.

TABLE 5

MODEL 1--1977 PREDICTION
(In thousands of Dollars)

Month	Prediction	Month	Prediction	Month	Prediction
Jan	8,802.3	May	8,928.2	Sep	9,054.1
Feb	8,833.8	Jun	8,959.7	Oct	9,085.5
Mar	8,865.3	Jul	8,991.1	Nov	9,117.0
Apr	8,896.7	Aug	9,022.6	Dec	9,148.5
Total					107,704.8

Model 2

As pointed out earlier, the residuals for Model 1 are correlated with one another. This is a common problem with time series data and is referred to as autocorrelation or serial correlation. A common method of remedying this problem is the use of autoregression, which employs "regression analysis to predict an independent variable when the dependent variables are merely lagged terms of the independent variable."[6] The general form of the autoregression equation is as follows:

$$Y_t = B_1 + B_2Y_{t-1} + B_3Y_{t-2} + \ldots + B_{n+1}Y_{t-n}$$

The next model to be developed is an autoregressive model which should not have correlated residuals as did Model 1.

The first autoregressive model is of a form which is widely used in Bell System forecasting of revenues. One will recall that MCR are highly trended and, being time series data, it is common to assume that expected MCR for next month are a function of what they are this month. It is also common to add to this assumption that MCR are also a function of what MCR were one year ago. These assumptions form the first attempt at an autoregressive model

$$M_t = b_1 + b_2M_{t-1} + b_3M_{t-12} + b_4M_{t-13} ,$$

where the lag of thirteen is added because of the interaction between the lags of one and twelve.

Using the same period as used in Model 1, January 1965 through July 1976, 13 data points are lost because of the thirteen lag period. Thus, the regression period for this model is February 1966 through July 1976 and n equals 126 data points.

Applying the least-squares regression equation, one obtains the following equation:

$$M_t = \underset{(45.748)}{159.43} + \underset{(.072716)}{.69054}M_{t-1} + \underset{(.098579)}{.26386}M_{t-12} + \underset{(.10157)}{.04258}M_{t-13}$$

Testing this model, R^2 is computed and found to be 99.7 percent. This measure of the "goodness of fit" tells one that 99.7 percent of the variation in the observed data is explained by the equation.

Testing the significance of the estimated coefficients, the following t-statistics are compared to the tabular value of ± 1.98:

TABLE 6

MODEL 2--COEFFICIENTS AND t-STATISTICS

Coefficient	Estimated Value	t-Statistic
b_1	159.43000	3.49
b_2	.69054	9.50
b_3	.26386	2.68
b_4	.04258	.42

Simply stated, in order to accept the null hypothesis that a coefficient is equal to zero, the computed t-statistic must be smaller than the tabular value. One can see that this is true for b_4. It must therefore be assumed that b_4 is equal to zero and is thus insignificant. For the other coefficients, the t-statistics are greater than the tabular value and are therefore assumed to be non-zero and significant.

Continuing with the tests of significance, the significance of the regression equation as a whole is tested with an F-test. For the estimated autoregression equation a computed F-statistic of 13,825.96 is obtained. This is compared to a critical value of approximately 2.68 for $F^{.05}_{3,122}$. The computed value is much larger than the tabular value, and one can assume that the estimated regression equation is significant with 95 percent confidence.

It is now time to perform a residual analysis which was the first hint of problems with Model 1. Figure 4 is a plot of the residuals for Model 2. A visual analysis of this plot is not fully conclusive; however, the residuals appear to be almost random except for a slight cyclical pattern. It can be said that the residuals in this plot appear to be more random than those for Model 1. This remaining pattern suggests that another variable should be

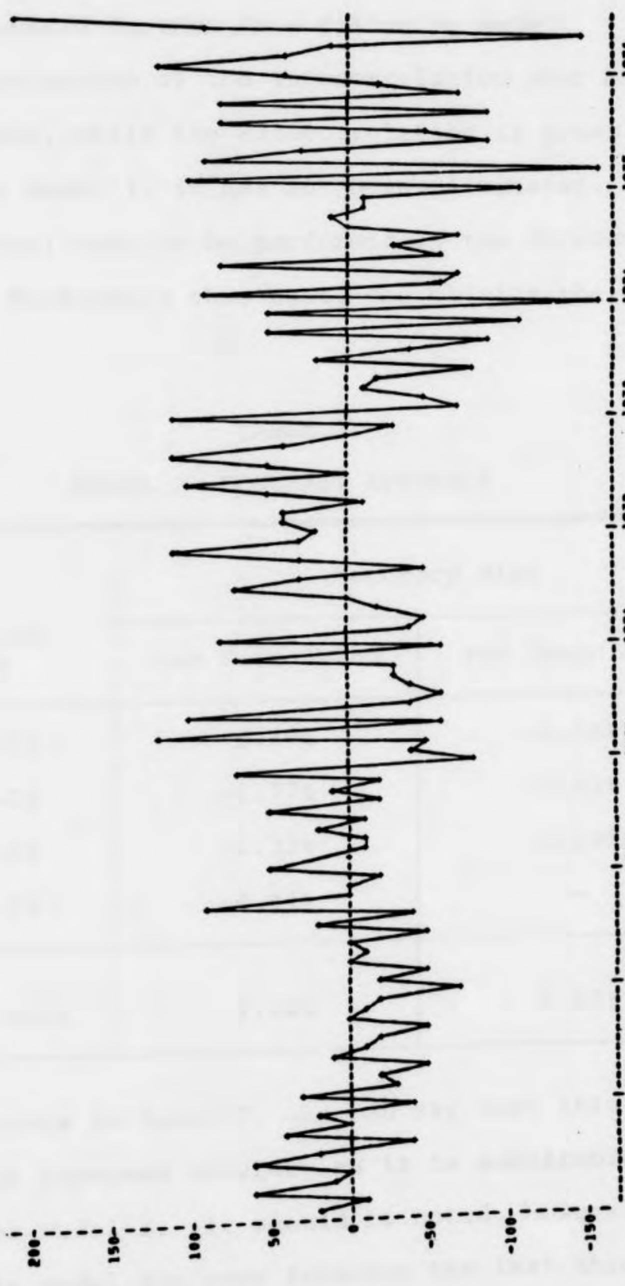

Fig. 4. Model 2--residual plot (February 1966 through July 1976).

added in order to explain the remaining cycle. This idea will be discussed further in a following model.

An examination of the autocorrelation plot in figure 5 tells one that, while the autocorrelation is greatly reduced over Model 1, it has not been eliminated.

The final test to be performed is the forecast accuracy test. Performing this test, one obtains the results in table 7.

TABLE 7

MODEL 2--FORECAST ACCURACY

Regression Period	Accuracy Miss	
	One Year Ahead	Two Years Ahead
Feb 1962-71	0.48%	-1.56%
Feb 1963-72	-1.77%	-3.81%
Feb 1964-73	-1.33%	-2.28%
Feb 1965-74	-0.44%	--
Average Absolute Miss	1.00%	2.55%

From the figures in table 7, one can say that this model also has poor forecast accuracy as it is comparable to the miss data for Model 1. It should be noted, however, that although this model has over forecast the last three years,

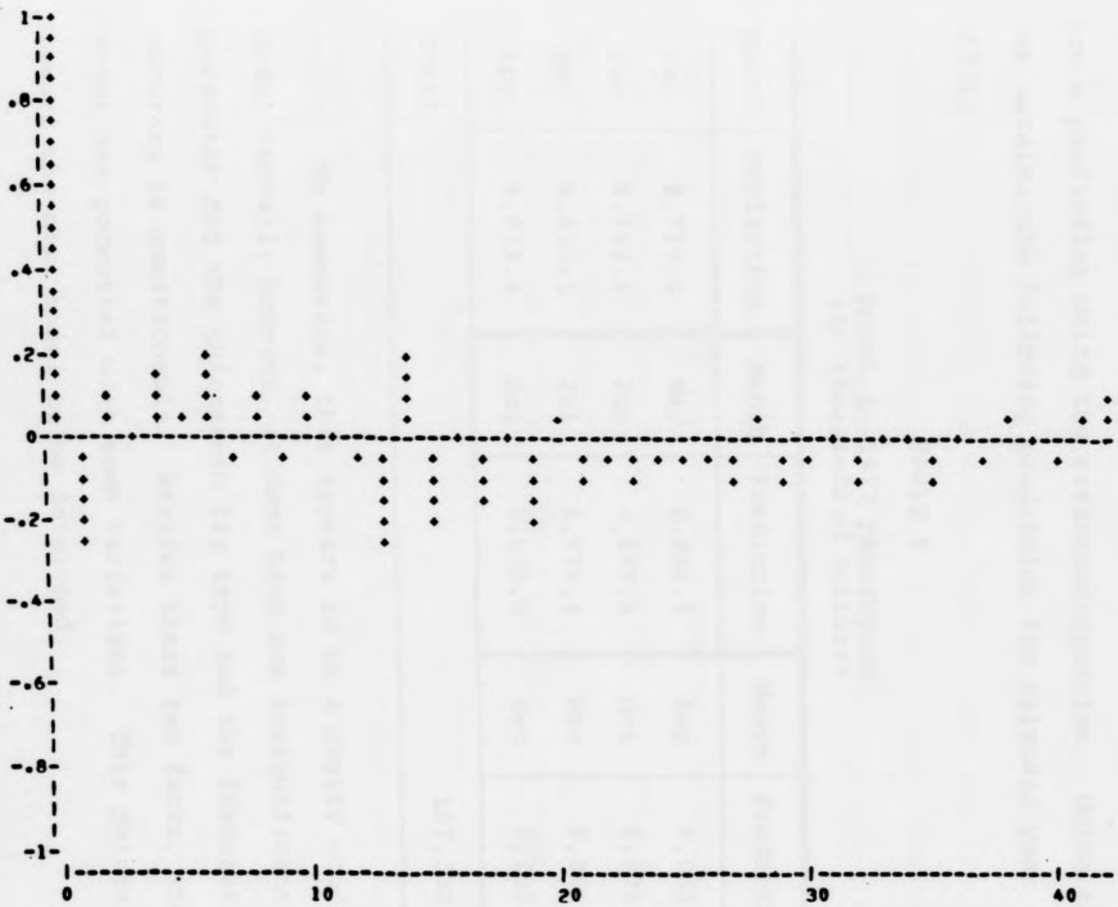

Fig. 5. Model 2--autocorrelation of residuals.

the amount of the miss is improving and suggests that it might be turning to an under forecasting position.

Having evaluated the model, it is now time to generate a prediction using the estimated equation. Doing this, one obtains the following prediction for calendar year 1977:

TABLE 8

MODEL 2--1977 PREDICTION
(In thousands of dollars)

Month	Prediction	Month	Prediction	Month	Prediction
Jan	8,719.0	May	8,869.8	Sep	9,061.6
Feb	8,782.5	Jun	8,897.9	Oct	9,079.8
Mar	8,836.1	Jul	8,979.4	Nov	9,092.5
Apr	8,875.4	Aug	9,035.3	Dec	9,103.1
Total					107,332.5

To summarize, this appears to be a pretty good model overall; however, it does have one insignificant parameter for the thirteenth lag term and the forecast accuracy is questionable. Besides these two facts, the model has potential with some variations. This points toward the next model to be developed.

Model 3

Autoregressive models can often be improved by first performing some level of differencing on the original time series data. It was discovered in Model 2 that lag parameters of one and twelve are significant paramaters,but the thirteenth lag parameter is insignificant. This would suggest that some differencing of the data might be in order. Since the thirteenth lag parameter is insignificant, a twelfth difference of the data is made. The lags of twelve and thirteen are dropped, leaving a lag of one. When the data are twelfth differenced, the original data are transformed to a series of twelve-month moving totals. The process removes the seasonality and "spurious correlation" from the data.[7] With the aid of a computer, this process is an easy one for both differencing the original data and undifferencing the predictions.

The new autoregressive model takes on the form

$$M_t^{12} = b_1 + b_2 M_{t-1}^{12} ,$$

where M^{12} signifies that the original data have been twelfth differenced.

Computing the least-squares regression, the following estimated equation is obtained:

$$M_t^{12} = \underset{(29.121)}{176.54} + \underset{(.07621)}{.53442} M_{t-1}^{12}$$

Testing the equation, R^2 is found to be 26.6 percent, which is low, much lower than has been experienced for the previous models. However, this measure is lower, because differenced data are being used, and should not be compared with the previous models.

Evaluating the significance of the estimated parameter coefficients, one compares the following t-statistics to the tabular value of ±1.98:

TABLE 9

MODEL 3--COEFFICIENTS AND t-STATISTICS

Coefficient	Estimated Value	t-Statistic
b_1	176.54000	6.06
b_2	.53442	7.01

Again, in order to accept the null hypothesis that a coefficient is equal to zero and insignificant, the computed t-statistic must be less than ±1.98. As one can see, this is true of neither of the coefficients; therefore, one must reject the null hypothesis for both coefficients.

Computing the F-statistic and comparing it to $F^{.05}_{1,136}$ one finds that the computed value, 49.18, is larger than the tabular value of 3.92. This means that the equation as a whole is significant.

Looking at a plot of the residuals in figure 6, they appear to be slightly more random than those of Model 2. Still some cyclical pattern remains, but the autocorrelation plot in figure 7 gives one satisfaction that the

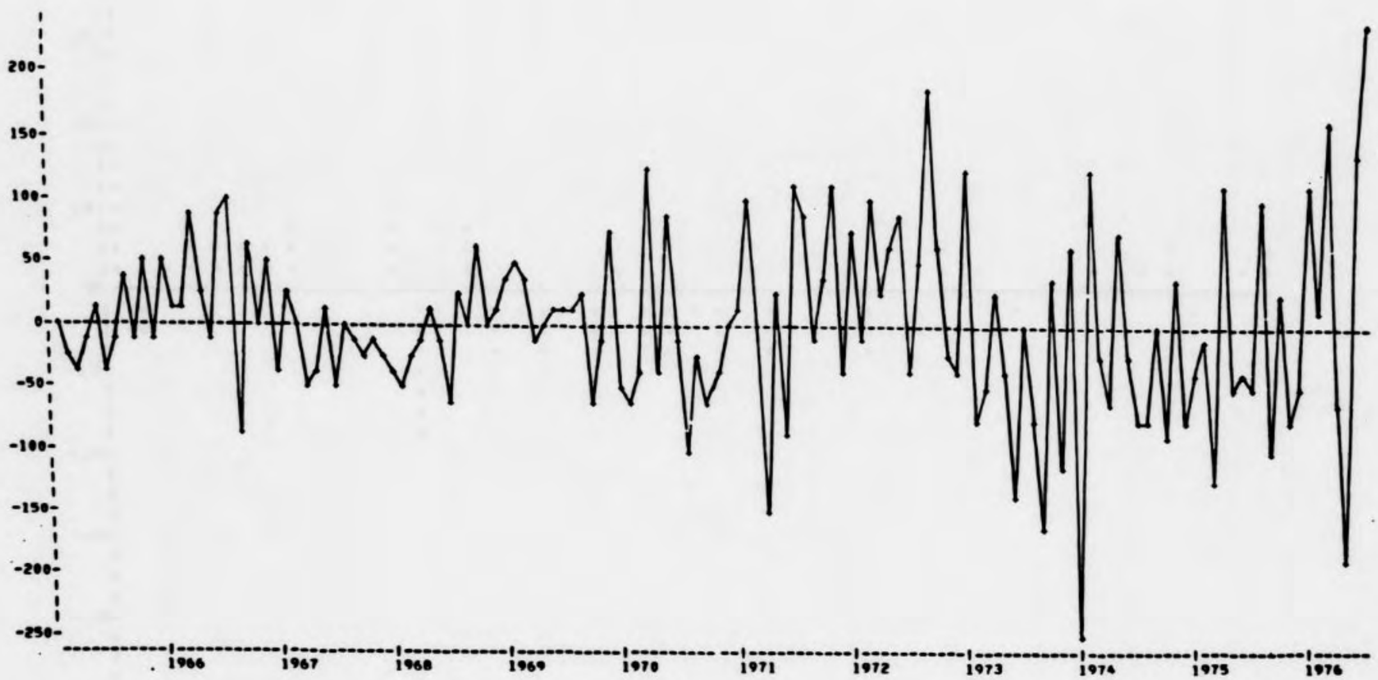

Fig. 6. Model 3--residual plot (February 1965 through July 1976).

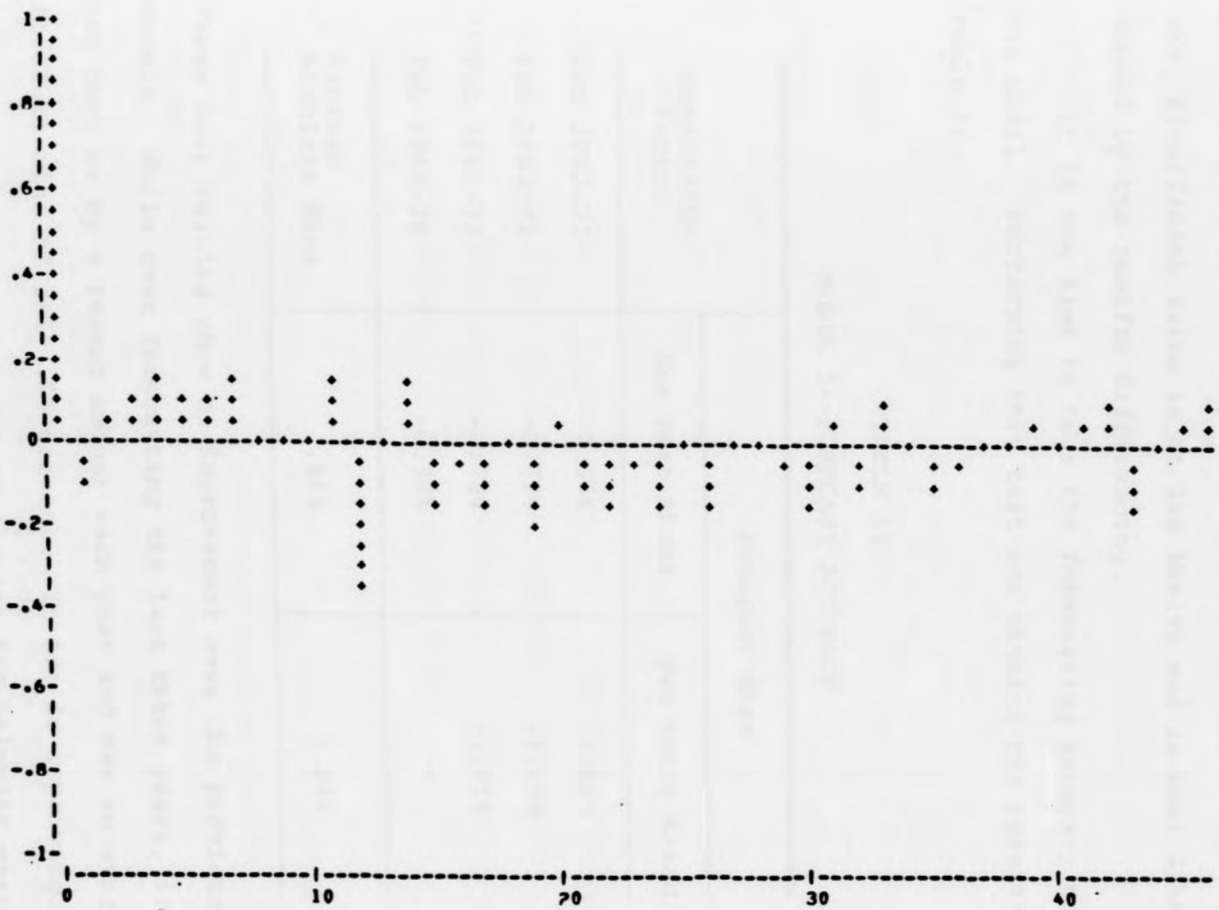

Fig. 7. Model 3--autocorrelation of residuals.

problem of autocorrelation has been remedied. The plot shows no pattern and only one significant value. However, the significant value is at lag twelve and is most likely caused by the twelfth differencing.

It is now time to test the forecasting accuracy of the model. Performing this test one obtains the results in table 10.

TABLE 10

MODEL 3--FORECAST ACCURACY

Regression Period	Accuracy Miss	
	One Year Ahead	Two Years Ahead
Feb 1962-71	1.57%	1.06%
Feb 1963-72	-0.78%	-1.46%
Feb 1964-73	-0.73%	-1.21%
Feb 1965-74	-0.38%	--
Average Absolute Miss	.86%	1.24%

These test results show an improvement over the previous models. While over forecasting the last three years, it has done so by a lesser amount each year and has an average absolute miss of only .63 percent. It too is improving.

Model 3 generates the prediction for calendar year 1977 set out in table 11.

TABLE 11

MODEL 3--1977 PREDICTION
(In thousands of dollars)

Month	Prediction	Month	Prediction	Month	Prediction
Jan	8,727.5	May	8,786.7	Sep	8,982.0
Feb	8,898.6	Jun	8,927.5	Oct	9,091.9
Mar	8,906.0	Jul	9,140.3	Nov	9,028.6
Apr	8,912.1	Aug	9,129.7	Dec	9,053.2
Total					107,584.1

In summary, this model appears to be a good one. It has passed all the statistical tests and has forecast fairly accurately. However, the model does have one fault, and that is the problem with the residuals. As was noted earlier, the residuals appear to have some cyclical pattern. This problem leads to the next model.

Model 4

As was mentioned earlier, the evidence of cyclical pattern in the residuals implies that another variable, one that would explain the cycle in the data, should be added to the model. That is the objective for Model 4. The variable that is added is Total Telephones In Service, as it stands to reason that MCR are a function of the number of telephones that generate the revenue. Actually the twelfth difference of Average Total Telephones In Service is added to the best autoregressive model, Model 3. The average in-service data are used because telephones are installed throughout the month and revenues are received for the portion of the month the telephone is in service. The lead/lag relationship is coincident, as revenues are booked in the month they are earned. Table 12 contains the actual values of Average Total Telephones In Service for the Period January 1960 through July 1976, of which February 1965 through July 1976 will be used in the model.

The new model takes on the following form:

$$M_t^{12} = b_1 + b_2 M_{t-1}^{12} + b_3 AT^{12} .$$

As one recalls, the first two parameters are taken from Model 3, where b_1 is the constant value and b_2 is the coefficnent for the lag of one. The additional parameter, b_3, represents the coefficient for the twelfth difference of Average Total Telephones In Service, represented here by AT^{12}.

Computing the least-squares regression, the following estimated equation is obtained:

$$M_t^{12} = \underset{(30.363)}{157.72} + \underset{(.7989)}{.48257} M_{t-1}^{12} + \underset{(.81759)}{1.6093} AT^{12} .$$

Evaluating the significance of the estimated parameter coefficients, one compares the following t-statistics to

TABLE 12

AVERAGE TOTAL TELEPHONES IN SERVICE
(In thousands)

Month	1960	1961	1962	1963	1964	1965	1966	1967	1968
	Year								
Jan	627.4	651.5	669.0	689.4	713.5	733.8	766.4	796.1	825.2
Feb	629.5	653.4	671.1	692.7	714.3	736.1	769.0	798.1	826.9
Mar	632.0	655.7	671.8	695.5	716.0	740.2	772.4	800.7	830.1
Apr	634.3	658.4	671.4	697.9	717.5	744.2	776.2	803.1	832.0
May	635.8	660.8	672.0	699.6	718.3	746.6	778.7	804.8	831.9
Jun	636.7	662.4	673.3	700.8	718.6	747.9	780.4	806.5	832.7
Jul	638.1	663.4	675.0	701.0	719.4	749.0	781.9	807.9	833.2
Aug	640.2	664.2	676.7	701.4	720.5	751.0	783.7	809.7	833.1
Sep	642.6	664.8	678.9	703.4	722.4	754.3	786.7	812.7	835.9
Oct	645.7	665.3	681.6	706.3	725.5	758.3	790.6	817.1	840.4
Nov	648.1	666.0	684.4	709.3	728.3	761.7	793.3	821.0	842.3
Dec	649.8	667.3	686.6	712.0	731.2	764.1	794.9	823.4	843.4

Month	1969	1970	1971	1972	1973	1974	1975	1976	
	Year								
Jan	845.6	871.2	900.9	936.4	957.6	971.9	981.1	993.4	
Feb	846.9	873.3	904.3	938.9	960.5	973.9	981.0	995.4	
Mar	847.8	875.6	907.6	941.8	963.2	975.6	984.3	996.8	
Apr	849.8	878.1	911.7	944.1	965.0	977.4	988.7	999.2	
May	852.1	880.0	914.3	943.3	964.6	977.6	988.2	1,000.0	
Jun	852.9	881.2	915.6	942.9	964.5	977.4	987.3	999.8	
Jul	853.8	882.5	915.7	944.3	965.0	976.7	988.6	1,000.8	
Aug	855.6	884.4	916.4	945.2	963.4	975.8	987.2		
Sep	858.5	888.3	920.6	948.2	964.9	977.9	987.8		
Oct	862.3	893.2	926.1	951.4	968.9	980.2	991.9		
Nov	866.1	895.8	930.7	953.2	970.5	980.9	994.4		
Dec	868.1	897.6	934.0	955.3	970.8	981.8	993.6		

the tabular value of 1.66. The tabular value of 1.66 is used because it can be assumed that the parameters will have positive values.

TABLE 13

MODEL 4--COEFFICIENTS AND t-STATISTICS

Coefficient	Estimated Value	t-Statistic
b_1	157.72000	5.19
b_2	.48257	6.04
b_3	.16093	1.97

Testing the null hypothesis that a coefficient is equal to zero, the t-statistic must be less than 1.66. This is true of none of the coefficients; therefore, one must reject the null hypothesis for all of the coefficients.

Analyzing the equation as a whole, the F-statistic is computed to be 27.05. This computed value is compared to the tabular value for $F^{.05}_{2,138}$ which is 3.07. Since the computed value is greater than the tabular value, one must assume that the estimated equation is significant.

Looking at the residual plot in figure 8, one cannot see much difference or improvement over the residuals for Model 3. The residuals appear to be almost random; however, the presence of some pattern still exists. The autocorrelation plot in figure 9 does suggest that no

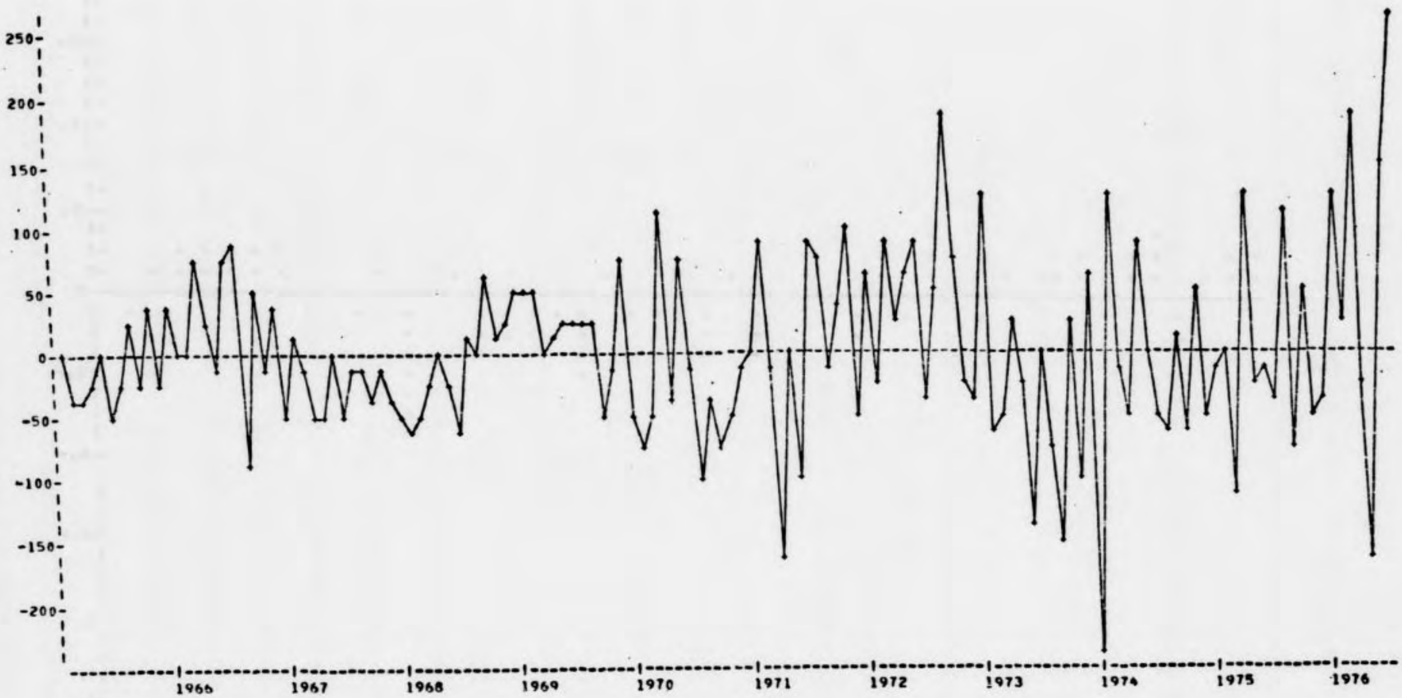

Fig. 8. Model 4--residual plot (February 1965 through July 1976).

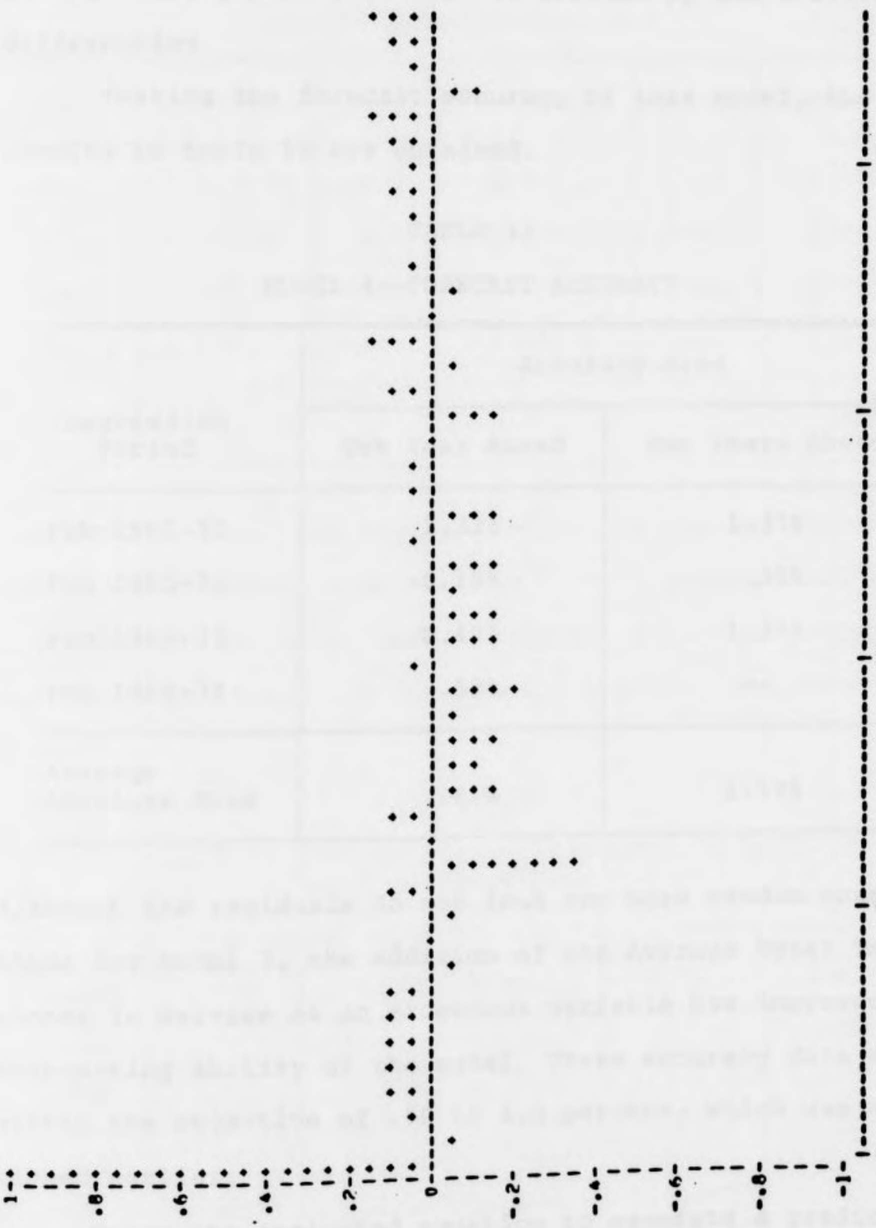

Fig. 9. Model 4--autocorrelation of residuals.

autocorrelation is present. There is a significant value at lag twelve, but again this is created by the twelfth differencing.

Testing the forecast accuracy of this model, the results in table 14 are obtained.

TABLE 14

MODEL 4--FORECAST ACCURACY

Regression Period	Accuracy Miss	
	One Year Ahead	Two Years Ahead
Feb 1962-71	1.32%	1.37%
Feb 1963-72	-0.16%	.60%
Feb 1964-73	0.40%	1.35%
Feb 1965-74	.59%	--
Average Absolute Miss	.61%	1.10%

Although the residuals do not look any more random than those for Model 3, the addition of the Average Total Telephones In Service as an exogenous variable has improved the forecasting ability of the model. These accuracy data are within the objective of .50 to 1.0 percent, which was mentioned earlier.

Using the estimated equation to generate a prediction, one obtains the information contained in table 15.

TABLE 15

MODEL 4--1977 PREDICTION
(In thousands of dollars)

Month	Prediction	Month	Prediction	Month	Prediction
Jan	8,710.8	May	8,779.9	Sep	8,942.3
Feb	8,886.0	Jun	8,920.4	Oct	9,053.6
Mar	8,897.3	Jul	9,133.2	Nov	8,992.9
Apr	8,905.6	Aug	9,093.3	Dec	9,023.6
Total					107,339.0

Table 16 contains the forecast values of Average Total Telephones In Service which were used to generate the above prediction. These estimates represent official forecasts that were generated for the company Construction Budget View.

TABLE 16

FORECAST OF AVERAGE TOTAL TELEPHONES IN SERVICE
(In thousands of dollars)

	Prediction			Prediction	
Month	1976	1977	Month	1976	1977
Jan	-	1,014.3	Jul	-	1,022.4
Feb	-	1,016.7	Aug	1,001.7	1,023.3
Mar	-	1,019.2	Sep	1,004.8	1,026.5
Apr	-	1,021.7	Oct	1,008.7	1,030.4
May	-	1,021.7	Nov	1,011.0	1,032.9
Jun	-	1,021.4	Dec	1,012.5	1,034.4

In summary, the model meets all the statistical tests and has good forecast accuracy. The model appears to be an excellent model.

Model 5

To this point two good models have been developed. Models 3 and 4 appear to be very sound. They pass all statistical tests and have performed well over the last few years in generating forecasts. However, one should not be satisfied to stop at this point. It is a sound forecasting practice to follow a multiple technique approach. If one can generate a number of models that produce forecasts that group within a rather close range, then a final forecast can be made with a greater degree of confidence. For this reason the modeling process will be continued and another model will be developed using a different technique.

The technique that is used now is not new in terms of invention but it is relatively new to forecasting practice. The process was introduced several years ago by two men, G. E. P. Box and G. M. Jenkins, and is often referred to as Box-Jenkins modeling. The process involves the use of two basic types of parameters, namely autoregressive parameters, which have already been used, and moving average parameters. The process also uses differencing, which has also been used in some prior models. The technique more simply involves a sophisticated smoothing process.

The idea is to employ a combination of differencing, autoregressive parameters and moving average parameters and try to make the data stationary. Data are said to be stationary when they have a constant mean and variance. Once the data are made stationary, least-squares regression is applied to that data. As was mentioned, the process is not new; however, before the days of the modern computer the manual process was so laborious that it was not a practical forecasting tool. Not only are complex manipulations required for the original data, but after the forecast is made those same manipulations must be reversed. Fortunately, today is known as the "age of the computer." Data can be manipulated with a machine, which, of course, is much faster, and it reduces the likelihood of human error.

The Box-Jenkins model building process[8] employs the use of two types of autocorrelation plots for identifying the differencing and parameters required. This is accomplished by identifying both significant autocorrelation values and meaningful patterns on the plots. In figure 10 is an autocorrelation plot for original base adjusted MCR. The vertical axis measures the autocorrelation values, while the horizontal axis represents a number of lag periods. The lines running parallel to the zero autocorrelation line represent 95 percent confidence limits outside of which autocorrelation is considered to be significant. This plot is dominated by an abundance of significant

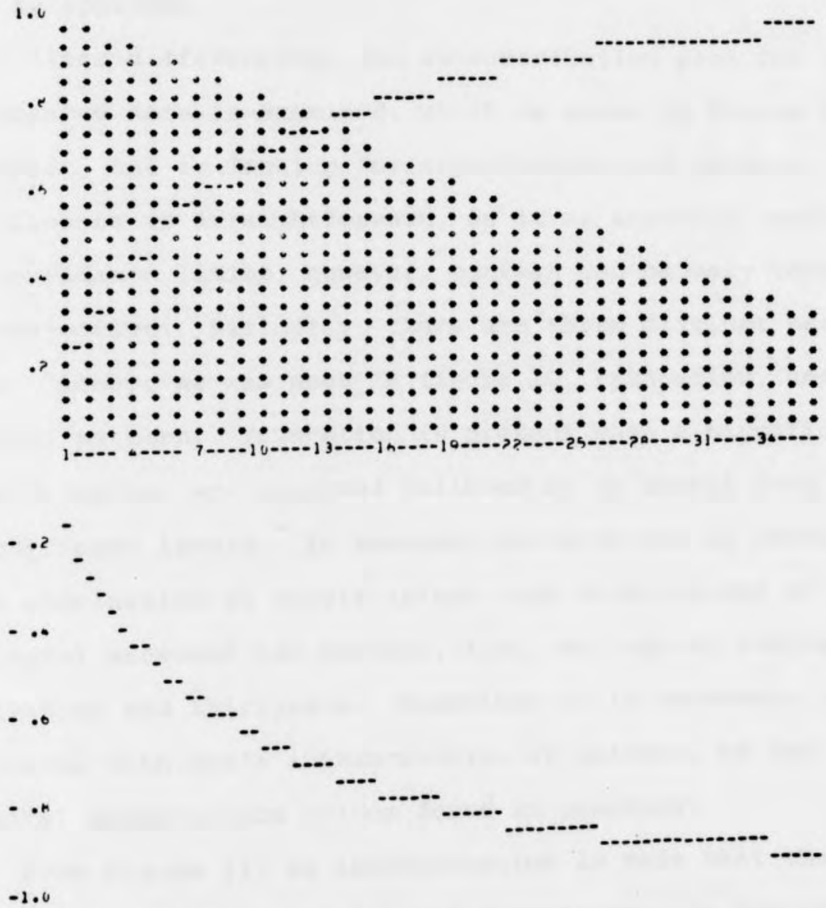

Fig. 10. Autocorrelation of MCR.

autocorrelation values with a decaying trend. These characteristics suggest that the data are dominated by trend, which, of course, is already known. In order to eliminate the trend component, a first difference of the original data is applied.

After differencing, the autocorrelation plot for the differenced data is examined, which is shown in figure 11. To repeat, one is looking for significance and pattern. Significance is straightforward, as it is anything outside the confidence limits; however, pattern can be very tenuous and subjective. Basically, there are three distinct patterns: decay, as was seen in figure 10, truncation, and seasonal pattern. Truncation is present when a significant spike or spikes are observed followed by an abrupt drop to insignificant levels. In seasonal patterns one is looking for a combination of single spikes that might appear at meaningful seasonal lag periods, i.e., at lags of twelve, twenty-four and thirty-six. Sometimes it is necessary to be liberal with one's interpretation of pattern, as the classical examples are seldom found in practice.

From figure 11, an interpretation is made that there is a significant spike at one and truncation. As stated earlier, one needs to examine an additional autocorrelation plot, one of partial autocorrelation, in order to determine the proper parameter to add to the model. It is the

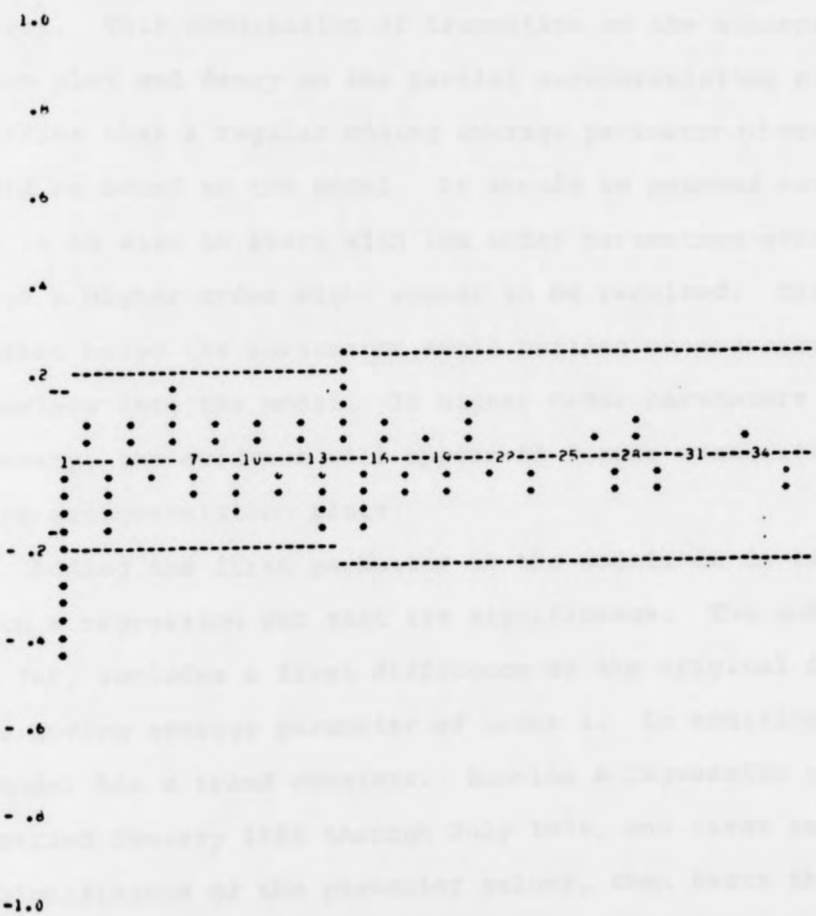

Fig. 11. Autocorrelation of first differenced MCR.

combinations of significance and pattern that are found on these two plots that tell what parameter to add.

Figure 12 is the partial autocorrelation plot for the data after applying a first difference to the original data. Here one sees significant values with a decaying pattern. This combination of truncation on the autocorrelation plot and decay on the partial autocorrelation plot signifies that a regular moving average parameter of order 1 should be added to the model. It should be pointed out that it is wise to start with low order parameters even though a higher order might appear to be required. This practice helps the forecaster avoid putting unnecessary parameters into the model. If higher order parameters are necessary, the evidence will appear in future examinations of the autocorrelation plots.

Adding the first parameter to the model, it is time to run a regression and test its significance. The model, thus far, includes a first difference of the original data and a moving average parameter of order 1. In addition, the model has a trend constant. Running a regression over the period January 1965 through July 1976, one first tests the significance of the parameter values, then tests that the residuals are in fact stationary.

A different approach than was used in the earlier models is used to test the significance of the parameter values. This is done because of a difference in computer

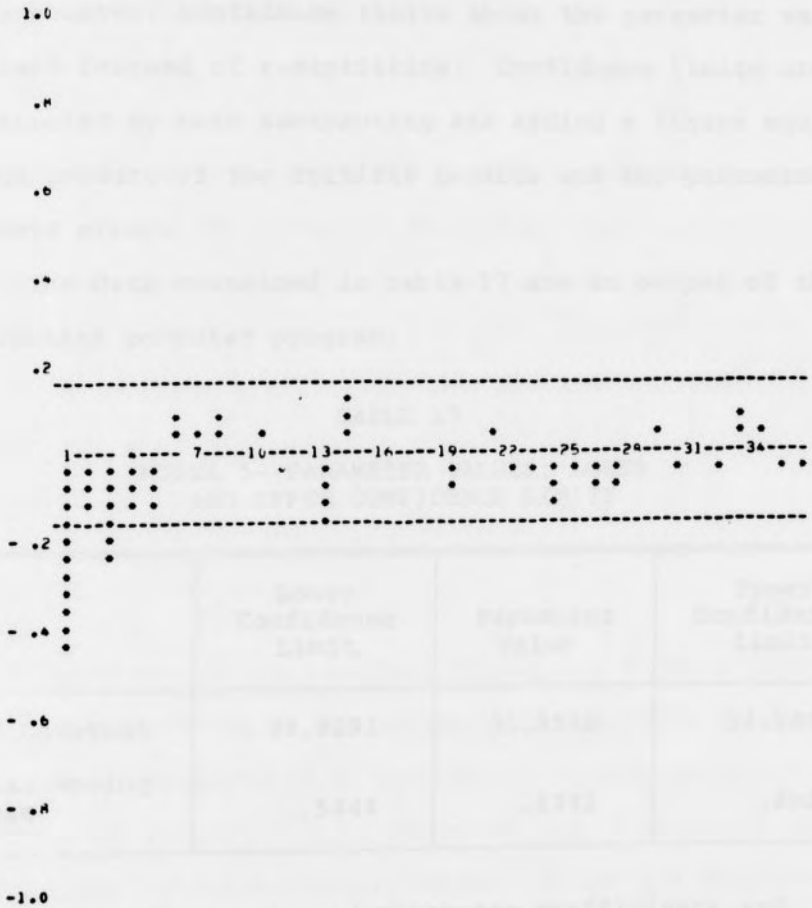

Fig. 12. Partial autocorrelation of first differenced MCR.

software and not because of necessity. As one recalls, in testing the significance of parameters in previous models, t-statistics were used in order to test the hypothesis that the values are not zero. The same hypothesis is tested here; however, confidence limits about the parameter values are used instead of t-statistics. Confidence limits are constructed by both subtracting and adding a figure equal to the product of the critical t-value and the parameter's standard error.

The data contained in table 17 are an output of the Box-Jenkins computer program.

TABLE 17

MODEL 5--PARAMETER VALUES, LOWER AND UPPER CONFIDENCE LIMITS

	Lower Confidence Limit	Parameter Value	Upper Confidence Limit
Trend Constant	28.9291	31.9590	34.9888
Regular Moving Average	.5444	.6732	.8020

The above are the estimated parameter coefficients and their lower and upper 95 percent confidence limits. Examining these data, one sees that the interval between the lower and upper confidence limits does not contain zero. This means that the hypothesis that either of these parameters is equal to zero can be rejected with 95 percent confidence. In other words, they are both significant.

In order to test the residuals for stationarity, one must test their variance with a chi-square statistic. The hypothesis to be tested is as follows:

$$H_0 : S^2 = 0 ,$$

that is, that the variance of the model residuals is equal to zero. If one can conclude that their variance is zero, then one can also conclude that stationarity has been obtained.

The equation for testing the mean and variance is as follows:

$$\chi^2(N) = \frac{\sum_i^N Y_i - \bar{Y}^2}{S^2} ,$$

where Y_i represents the residual values and $\bar{Y}$ their mean. Such a chi-square value for the model's residuals is 29.81. This value is compared to a tabular chi-square value with 48 degrees of freedom. This value for the 95 percent confidence level is approximately 67.50. Since the computed value is less than the tabular value, one can accept the hypothesis that the variance of the residuals is equal to zero and stationarity has been obtained.

For further evidence of the effectiveness of the model, one can examine the autocorrelation and partial autocorrelation plots of the residuals. The autocorrelation plot is contained in figure 13. Here one sees that the residuals appear to be random. There is one spike outside the confidence limits at a lag of 14; however, there is nothing meaningful about a lag 14, and there is a 5 percent chance that a value will appear outside the confidence limits and be insignificant.

Examining the partial autocorrelation plot in figure 14, nothing of significance is present. Again there is a spike at 14, but it is concluded not to be meaningful, because there is no conceivable reason for a lag of 14 to be significant.

It is now time to test the forecasting ability of this model. Testing the model in the same manner as the previous models, the results in table 18 are obtained.

TABLE 18

MODEL 5--FORECAST ACCURACY

Regression Period	Accuracy Miss	
	One Year Ahead	Two Years Ahead
Jan 1962-71	0.98%	0.46%
Jan 1963-72	-0.95%	-1.67%
Jan 1964-73	-0.77%	-1.27%
Jan 1965-74	-0.50%	--
Average Absolute Miss	.80%	1.13%

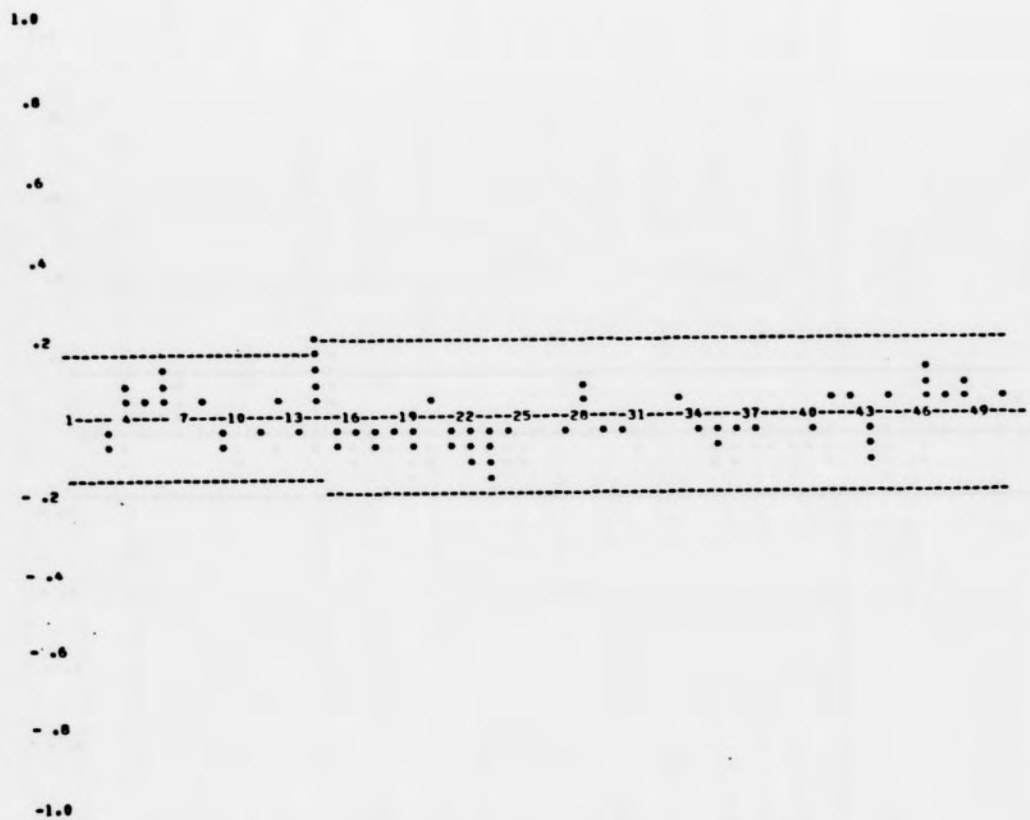

Fig. 13. Model 5--autocorrelation of residuals.

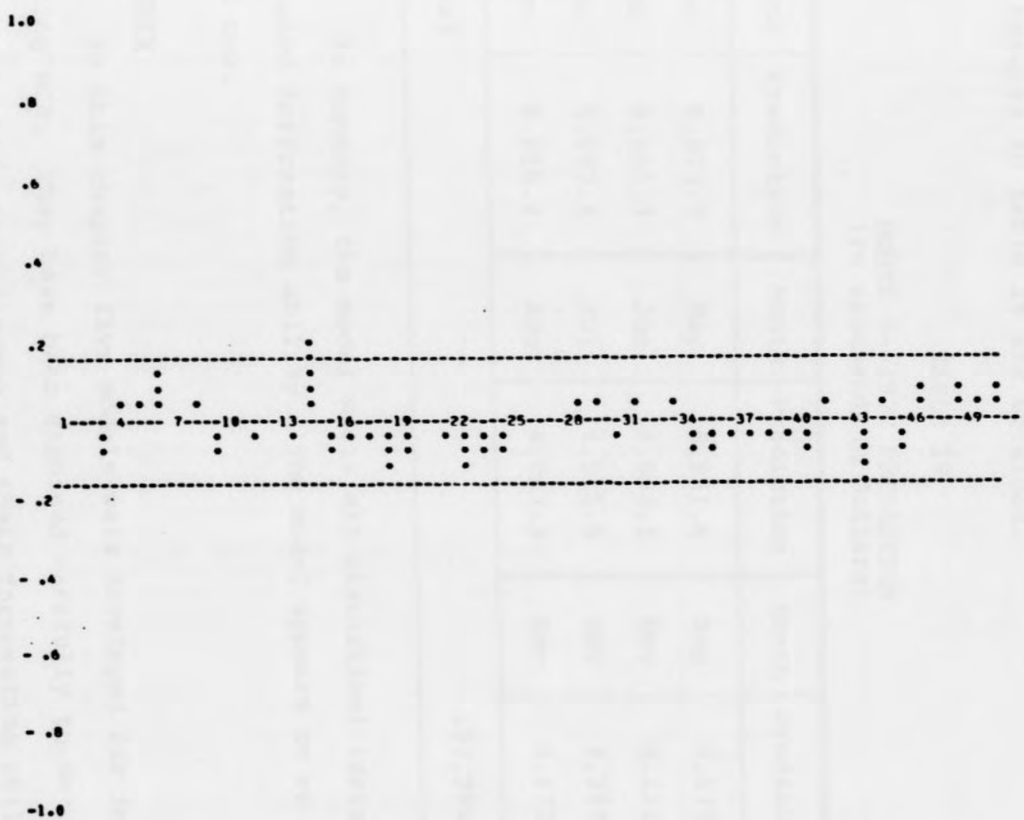

Fig. 14. Model 5--partial autocorrelation of residuals.

The miss data in table 18 appear to be reasonably good and compare favorably with some of the previous models.

Using the model to generate a prediction for 1977, the results in table 19 are obtained.

TABLE 19

MODEL 5--1977 PREDICTION
(In thousands of dollars)

Month	Prediction	Month	Prediction	Month	Prediction
Jan	8,823.7	May	8,951.6	Sep	9,079.4
Feb	8,855.7	Jun	8,983.5	Oct	9,111.4
Mar	8,887.6	Jul	9,015.5	Nov	9,143.3
Apr	8,919.6	Aug	9,047.4	Dec	9,175.3
Total					107,994.0

In summary, the model meets all statistical tests and has good forecasting ability. The model appears to be a good one.

Summary

In this chapter five models were developed for forecasting MCR. They have been examined carefully to determine both their significance and their forecasting ability. The five models have also been used to generate forecasts of MCR for calendar year 1977. These models are summarized in table 20, where a quick review of the models, their

TABLE 20

MODEL SUMMARY

Model No.	Model Definition	R^2	F	X^2	Accuracy Miss 1 Yr Ahead	Accuracy Miss 2 Yrs Ahead	1977 Prediction (in $'000)
1	$M_t = \underset{(3360.9)}{-73732} + \underset{(.14219)}{31.466T_t}$ Comments: Autocorrelation present, tests of significance questionable. Poor forecast accuracy.	.997	48,973	NA	1.08%	1.22%	107,705
2	$M_t = \underset{(45.748)}{159.43} + \underset{(.072716)}{.69045M_{t-1}} + \underset{(.098579)}{.26386M_{t-12}} + \underset{(.10157)}{.04258M_{t-13}}$ Comments: Coefficient for M_{t-13} insignificant. Autocorrelation present. Poor forecast accuracy.	.997	13,826	NA	1.00%	2.55%	107,333
3	$M_t^{12} = \underset{(29.121)}{176.54} + \underset{(.07621)}{.53442M_{t-1}^{12}}$ Comments: Passes all tests of significance. Good forecast accuracy. Residuals have a cyclical pattern.	.266	49.18	NA	.86%	1.24%	107,584
4	$M_t^{12} = \underset{(30.363)}{157.72} + \underset{(.07989)}{.48257M_{t-1}^{12}} + \underset{(.81759)}{1.6093AT^{12}}$ Comments: An excellent model. Has had a tendency to under forecast.	.286	27.05	NA	.61%	1.10%	107,339
5	$M_t = 31.959 + A_t - .6732A_{t-1}$ Comments: A good model. Improving forecast accuracy. Has over forecast the last three years.	NA	NA	29.81	.80%	1.13%	107,994

test statistics and a forecast using the model is reviewed.

The first model developed was a regular linear regression of MCR on time. All the test statistics suggested that the model is a good one; however, autocorrelation was found to be present. When autocorrelation is present, the test statistics are unreliable. The model also failed to meet the standards established for forecast accuracy. Overall the model must be rated as very poor.

Model 2 is an autoregressive model with lags of 1, 12 and 13 months. With exception to the coefficient for the lag of 13, all tests suggested that the model is a good one. Some autocorrelation was present, and the model's forecasting ability was also poor. Like the first model, Model 2 must also be rated poor.

Model 3 is a modified version of Model 2. It too is an autoregressive model. It was stated above that the coefficient for the lag of 13 proved to be insignificant. Instead of simply dropping this parameter from the model and estimating the equation over, a different modification was made. A twelfth difference was taken of original MCR, and the lags of 12 and 13 were dropped from the model. These modifications proved to be very beneficial. The model passed all tests of significance, and the problem of autocorrelation was remedied. The model's ability to forecast the last four years was also greatly improved. This model is rated good.

Model 3 did have one minor flaw. The residuals appeared to have a cyclical pattern. In an attempt to remedy this flaw, an exogenous variable was added to the model. MCR were assumed to be a function of the number of telephones in service; therefore, a series for "Average Total Telephones In Service" was added to Model 3. The results proved to be very significant. The model passed all tests of significance. Although it did not correct the cyclical pattern in the residuals, it did reduce the absolute average forecast miss to .61 percent for one year ahead and 1.10 percent for two years ahead. Model 4 is clearly the best model developed to this point. It should be pointed out that the model has tended to under forecast.

The final model developed is an example of a sophisticated smoothing technique referred to as Box-Jenkins models. The model is a simple one with a trend constant and moving average parameter of order one. A first difference was also applied to the original data. The resulting model was a good one. Its forecast ability was adequate, but more importantly it was improving.

FOOTNOTES

[1]John G. Myers, "Statistical and Econometric Methods Used in Business Forecasting," Methods and Techniques of Business Forecasting, ed. William F. Butler, Robert A. Kavesh, and Robert B. Platt (Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1974), p. 12.

[2]All data manipulations, regression analysis, tests of significance, forecasts, etc., presented in this thesis are the output of programs of the Bell Labs Statistical Computing Library (STATLIB) and were generated through the use of the time share computer facilities of the National Computer Software System, Norwalk, Connecticut.

[3]Myers, "Statistical and Econometric Methods," p. 17.

[4]Ibid., pp. 17-19.

[5]Ibid., p. 20.

[6]William L. Hays and Robert L. Winkler, Statistics: Probability, Inference, and Decision, 2 vols. (New York: Holt, Rinehart and Winston, Inc., 1970), 2:80.

[7]For an explanation of "spurious correlation," the reader is referred to Hays and Winkler, Statistics, 2:80-81.

[8]For more detail on the process of Box-Jenkins modeling, the reader is referred to George E. Box and Gwilym M. Jenkins, Time Series Analysis Forecasting and Control (San Francisco: Holden-Day, 1970).

CHAPTER 5

SUMMARY AND CONCLUSIONS

It was stated in chapter 1 that the ultimate goal of this thesis is to make an adequate forecast of MCR for calendar year 1977. It is now time to make that forecast. However, first, a brief review of the work presented thus far is in order.

Summary

First of all, the nature of the Bell System and the system's commitment to provide the services demanded at a regulated price suggest that demand is the driving force for the determination of the company's revenue in the short run. For this reason, a review of the literature concerning the demand for telephone service was undertaken in chapter 2.

The literature reinforced the idea that Bell System demand is dependent upon external economic factors. Demand models were reviewed where the variables used consisted of measures of the national economy, demography, personal income, consumption, prices, consumer tastes, etc. However, it was determined that this was not the proper approach to follow in forecasting for a small area like the city of Washington, D.C. The lack of adequate data and the unique nature of the District's economy were instrumental in this conclusion.

The literature also suggested that forecasting accuracy could be improved by dividing the aggregate into subaggregates and modeling the subaggregates. This approach has proven to be successful in forecasting telephone demand where the aggregate, total telephone demand, has been divided into residence and business telephone demand. While this would be a practical approach to apply to forecasting revenues, it was impossible under current circumstances because revenues in the Bell System are not reported in such a subaggregate nature. For this reason, separate historical data series for residence- and business-generated revenues do not exist.

As alternatives to the use of economic variables and analysis of subaggregates, the literature recommended the use of autoregression and smoothing techniques.

In chapter 3, various tables and plots were used along with an analysis of variance in order to examine the underlying characteristics of MCR. An examination of a plot of the data visually showed that the data are dominated by trend. This fact was quantified with an analysis of variance on a two-way table of the data where the columns were identified by years and rows by months. Such an analysis showed that 99.5 percent of the variation in the data is explained by trend and cycle. That same analysis showed that .4 percent of the variation is explained by seasonality and .1 percent by irregularity.

In chapter 4, five models of MCR were developed. The models were tested for significance and forecasts were made. Of all the five models developed, Models 4 and 5 were the best and will now be used to make the forecast. Model 4 produced a forecast of $107,339,000; however, this model has had a tendency to under forecast the last four years. Because of this tendency to under forecast, it is appropriate to make an adjustment to this prediction. Over the period of 1972 through 1975, Model 4 under forecast one year ahead by an average of .54 percent. This miss is applied to the original forecast to produce a new forecast of $107,919,000.

Model 5 produced a forecast of $107,994,000 for 1977. Unlike Model 4, its tendency has been to over forecast. However, this tendency has been declining, and there is a strong possibility that it could under forecast 1977. This is evidenced by the fact that by applying the model to data through December 1975 and forecasting the first seven months of 1976, the model under forecast those seven months consistently and by 1.15 percent. Considering all of the above, a somewhat conservative forecast of $108,000,000 is selected for 1977.

In chapter 3 it was mentioned that C&P of Washington was granted a rate increase effective June 1, 1976. At the time the models were developed and the resulting forecast made, a lack of information and time made it necessary to ignore the rate increase until now. The lack of time and information made it impossible to adjust MCR in order to account for the rate increase. Under these circumstances, it is necessary to rely on an estimate of the value of the rate increase which is furnished by the Comptroller's organization. This estimate amounts to approximately \$3,700,000. Adding this figure to the forecast made from the models, the final forecast for 1977 is \$111,700,000.

Direction for Future Research

Having made the forecast for 1977, a few remarks concerning the direction of future investigations are appropriate. First of all, an analysis of the degree of accuracy that one should strive to attain should be made. AT&T suggests that .5 to 1.0 percent is adequate; however, the forecast user should be questioned to determine the degree of accuracy desired. Important questions should also be asked concerning the penalties of over and under forecasting. Considering the outcome of this investigation, decisions can be made about the need for further research of appropriate economic variables to be used in an econometric model. Depending on the need and benefits of improving the forecast, it can be determined if the time and computer expense are worth the effort.

Finally, steps should be taken to report company revenues in a variety of subaggregates. Revenues definitely should be divided between residence and business. An investigation should be made for suggestions of other such divisions.

BIBLIOGRAPHY

Books

Box, George E., and Jenkins, Gwilym M. Time Series Analysis Forecasting and Control. San Francisco: Holden-Day, 1970.

Chisholm, Roger K., and Whitaker, Gilbert R., Jr. Forecasting Methods. Homewood, Ill.: Richard D. Irwin, Inc., 1971.

Christ, Carl F. Econometric Models and Methods. New York: John Wiley & Sons, Inc., 1966.

Haitovsky, Yoel; Treyz, George; and Su, Vincent. Forecasts with Quarterly Macro-Econometric Models. New York: National Bureau of Economic Research, 1974.

Hays, William L., and Winkler, Robert L. Statistics: Probability, Inference, and Decision. 2 vols. New York: Holt, Rinehart and Winston, Inc., 1970.

Johnston, J. Econometric Methods. 2d ed. New York: McGraw-Hill Book Company, 1972.

Malinvand, E. Statistical Methods of Econometrics. Chicago: Rand McNally and Company, 1966.

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

Wolfe, Harry Deane. Business Forecasting Methods. New York: Holt, Rinehart and Winston, 1966.

Wonnacott, Ronald J., and Wonnacott, Thomas H. Econometrics. New York: John Wiley and Sons, Inc., 1970.

Articles

Chaddha, Roshan L., and Chitgopekar, Sharad S., "A Generalization of the Logistic Curves and Long-range Forecasts (1966-1991) of Residence Telephones." The Bell Journal of Economics and Management Science 2 (Autumn 1971):542-60.

Davis, B. E.; Caccappolo, G. J.; and Chaudry, M. A. "An Econometric Planning Model for American Telephone and Telegraph Company." The Bell Journal of Economics and Management Science 4 (Spring 1973):29-56.

Dunn, Douglas M.; Williams, William H.; and Spivey, Allen W. "Analysis and Prediction of Telephone Demand in Local Geographical Areas." The Bell Journal of Economics and Management Science 2 (Autumn 1971):561-76.

Evans, Michael K. "Econometric Models." In Methods and Techniques of Business Forecasting, pp. 161-89. Edited by William F. Butler, Robert A. Kavesh, and Robert B. Platt. Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1974.

Myers, John G. "Statistical and Econometric Methods Used in Business Forecasting." In Methods and Techniques of Business Forecasting, pp. 11-36. Edited by William F. Butler, Robert A. Kavesh, and Robert B. Platt. Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1974.

Rohlfs, J. "A Theory of Interdependent Demand for a Communications Service." The Bell Journal of Economics and Management Science 5 (Spring 1974):16-37.

Thompson, Howard E., and Tiao, George C. "Analysis of Telephone Data: A Case Study of Forecasting Seasonal Time Series." The Bell Journal of Economics and Management Science 2 (Autumn 1971):515-41.

von Rabenau, Burckhard, and Stahl, Konrad. "Dynamic Aspects of Public Goods: A Further Analysis of the Telephone System." The Bell Journal of Economics and Management Science 5 (Autumn 1974):651-69.