REGIONAL AND
INTERREGIONAL ASPECTS

REGIONAL ECONOMIC FORECASTING: CONCEPTS AND METHODOLOGY

W. ALLEN SPIVEY and WILLIAM E. WECKER
University of Michigan

Most forecasting models can be grouped into one of two classes: intrinsic (or extrapolative) and extrinsic (or associative). By intrinsic models we mean those in which one forecasts future movements of a variable on the basis of an analysis of past movements of the variable, temporarily ignoring the influence of other variables upon the variable of interest. These models usually deal with a time series of observations on the given variable; they may be simple models such as linear or "straight-line" models and the forecasting may be little more than naive extrapolation of a linear function of time that is fitted to the data by visual or least-squares procedures, or they may be more sophisticated models such as exponential smoothing (Brown [4]) or adaptive exponential smoothing (Trigg and Leach [32]; Dunn, Williams, and Spivey [12]). No matter how sophisticated the assimilation of past data into the model may be, in this model class the influence of related variables is essentially ignored.

In extrinsic or associative models, one attempts to relate the movements of a variable to movements in one or more related variables. One may do this through a regression model, through a system of econometric equations, or through a system of deterministic linear equations such as an input-output model.[1] Many of the forecasting models developed in recent years in the regional analysis literature have been of this class. Economic base studies are essentially associative models as are the econometric models of Anderson [1], Glickman [13], Bell [2], and Mattila [24].

Each of these model classes has its advantages and disadvantages. Intrinsic models are available in many kinds, they do not require a data base for related variables, and a sensitivity study of forecasting accuracy using different parameter values can be made fairly easily. They clearly suffer because they do not include influences from related variables, and the results of using them are sensitive to the way in which trend problems are dealt with in the data and in the model. The problem of trend in an economic time series is very complex and there is at present no clearly defined way of handling it well (the problem bedevils extrinsic models as well).

Extrinsic models appeal to many because they often appear to express or capture interrelationships that one's knowledge of economics suggests are important.

---

[1] This classification is virtually the same as that of Isard *et al.* [18]; intrinsic models essentially use direct techniques and extrinsic models are indirect techniques.

However, very large quantities of data are required, it is difficult to assess forecasting accuracy against alternative models of the same class, and problems of autoregression, aggregation, stochastic dependence, and multicollinearity are very difficult to deal with constructively. It is the view of the authors that data base problems alone are severely inhibiting in the use of these models for regional forecasting at the level of a state or a smaller region and that time-series methods emphasizing intrinsic models are worthy of additional study and experimentation.

It seems obvious that one should be able to combine features of intrinsic and extrinsic models in forecasting, depending on the problem one faces. Interestingly enough, few examples of doing this can be found in the economics and statistics literature. Such a model, which for lack of a better term we call a hybrid model, would have appealing features for regional forecasting since it would offer the hope, if only a small number of associated variables are necessary, of reducing the problem of data gathering.

The purpose of this paper is to present some of our experience in using various intrinsic models for regional forecasting, to indicate the nature of some of our supporting data analyses, to show how these studies influenced us in the model development, and, finally, to indicate some avenues for future research. As part of the latter we show the results we obtained with a hybrid model in which information from one associated variable was used. A comparison of the forecasting performance of this model with that of the others is given below and the reduction in forecast error, despite the tentative nature of the model, indicates that hybridization may well be a promising approach for future study.

The models appear in the form of a case study discussion below, which explains how one can proceed from one model to another. Two of the models and some of the supporting data analysis appear in Dunn [10] and in Dunn, Williams, and Spivey [12]; the remainder of this section has been developed independently by the authors.

We give only a brief discussion of each forecasting model used, citing the appropriate literature where additional detailed information can be obtained, and our approach is pragmatic: we look at the forecasting performance of each model and present a corresponding error analysis and, in most cases, some graphical output which shows the time series we are forecasting and the time series that is generated by the forecasting model.

Statement of the Forecasting Problem

We are addressing the problem of forecasting the growth in telephone demand in local areas of Michigan. By "local area" we mean a geographical area served by a telephone switching facility called a wire center. Such a facility can serve all or part of a county or all of or parts of two counties. Thus the usual problems of regional forecasting are present: there are grave data limitations, the region is not necessarily a political subdivision such as a city, county, or Standard Metropolitan

Statistical Area, and data on related variables are either scarce or nonexistent. More-over, the problem is an important operational one: the Bell System spends more than 2 billion dollars per year on new plant and equipment in wire centers (about $150 million per year in the region served by Michigan Bell alone) and these expen-diture decisions are based upon the forecasted growth of telephone demand in a
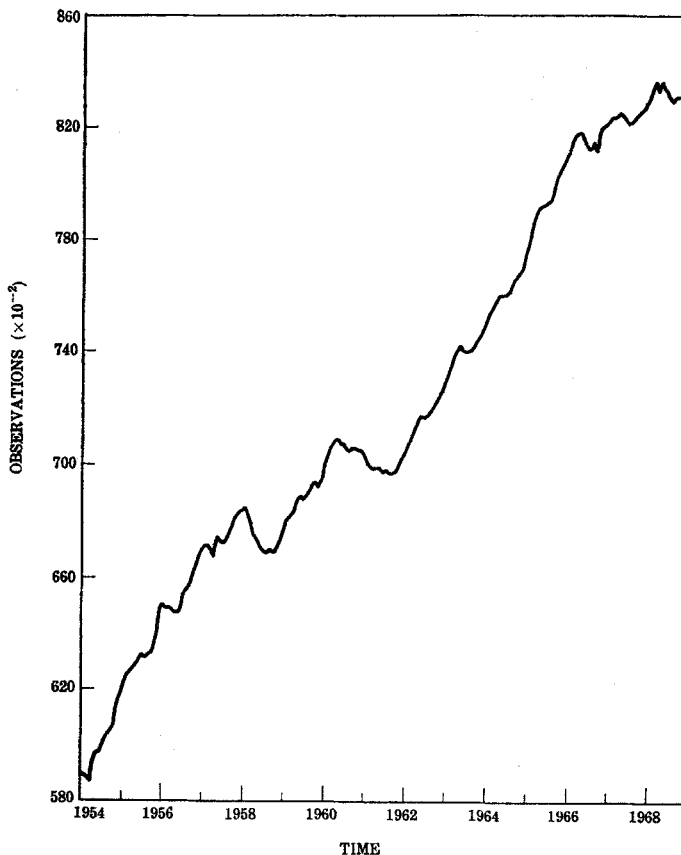

FIGURE 1. RESIDENCE MAIN STATIONS FOR FLINT

twelve- to eighteen-month time period into the future. Finally, there are more than 300 wire centers in the region served by Michigan Bell for which forecasts are needed on at least an annual basis, so a forecasting model cannot be costly or cumbersome to use. A detailed statement of this problem setting appears in Dunn, Williams, and Spivey [12]. The models and specific forecasts appearing below refer to forecasts developed for the Flint region, which includes all of Genessee county and parts of Lapeer county (the data are monthly from January, 1954, to December, 1969,

and are used with the permission of Michigan Bell Telephone Company).[2]

The time series is shown in Figure 1. It looks like many economic time series in that it is "trending upward" over the entire time period but its trend is extremely difficult if not impossible to define. The extrapolation of a straight line fitted to all the data, for example, is clearly inappropriate and the extrapolation of a straight line fitted to parts of the data is unwise also because of the horizontal "drifts" or displacements that occur (but not with regularity).

Data Analysis and Model Building

No one model is necessarily best for all forecasting problems. Models should be data-oriented and responsive to the peculiarities of the data one has to deal with. Sometimes sophisticated models lead one astray when based upon an inadequate study of the underlying data. The case for data analysis as an activity of equal importance with statistical analysis and inference is persuasively made by Tukey [34] and Tukey and Wilk [35].

A useful tool of data analysis is spectral analysis. In particular, the analysis of the sample spectrum or of the sample autocorrelation function of a time series may reveal periodicities, seasonal influences, or other features which may be exploited in the development of a forecasting model. A detailed discussion of how spectral analysis can be used in the course of developing a forecasting model can be found in Dunn [10] and in Granger and Hatanaka [15].

Sometimes data analysis suggests that a time series can or should be decomposed into two or more component time series, each of which can be analyzed and forecasted separately. The component forecasts can then be aggregated into a forecast of the parent series and this approach may be superior to any procedure used on the parent series alone.

Sometimes a decomposition can be suggested in a very simple way: a parent series, regarded as a random variable, may be the sum of two or more random variables, and it may be more productive to deal with the component random variables individually. Such a case occurred in the Flint telephone problem: the time series one is working with is the number of residential telephones (main stations) on a monthly basis (this is the time series shown in Figure 1).[3]

---

[2] It might be mentioned that early in the study of this problem regression analysis was used and a multiple regression equation was developed in which the independent variables were taken to be total covered employment, average hourly earnings, index of Michigan industrial activity, index of U.S. industrial activity, residential construction in square feet, and total dollar value of residential construction (all data except the indexes were for Genessee county). Very poor forecasting results were obtained by using regression procedures on these and other variables, so an intrinsic forecasting strategy was adopted.

[3] A residential main station is one or more instruments with the same telephone number. The switching capacity necessary to serve two or more instruments having the same number is about the same as that for serving one telephone, whereas two instruments with two numbers in the same location requires much more switching capacity. Hence a main telephone is the basic variable of interest. Moreover, for reasons set forth in Dunn, Williams, and Spivey [12], the time series of residential main stations is regarded as adequately representing residential telephone demand.

The change in the number of total telephones in a month is the sum of "connects" and "disconnects" in that month. It turned out that the latter two time series displayed different patterns over time so these series were analyzed separately, a model was developed for each, forecasts were generated, and these were combined to obtain a forecast of residential telephones. Handling the problem this way resulted in better forecasts than we were able to obtain by forecasting directly the parent series of residential telephones. (There are other and more sophisticated procedures for decomposing a time series; see, for example, Malinvaud [23]).

An important aspect of data analysis is the study of residuals from some chosen fitting function. One may fit a straight line or higher order polynomial to a time series, determine the "spread" of the data points around this function (the difference between the function value and the corresponding observation value is called a residual). It may be that one encounters more success with forecasting procedures applied to the residual series than to the original series. In any case, an analysis of the residuals is often essential in order to develop an adequate understanding of the original data.

Choosing a Forecasting Model

Before comparing the forecasting performance of several models, it is appropriate to comment on the criteria one is to use. Since a forecast can be regarded as a point estimate of $E(X_{t+\tau} | X_t, X_{t-1}, \ldots)$, where τ is the lead time of the forecast, the conventional statistical criteria of point estimation seem to be a natural choice. For example, one might prefer an estimator that is unbiased and has minimum variance among the class of unbiased estimators. If a loss function relating to forecast error can be determined, one might choose a point estimator which is optimal in the sense of minimizing the expected loss. Unfortunately, the above criteria are difficult if not impossible to apply for many forecasting models because the distribution of the forecast errors is unknown.
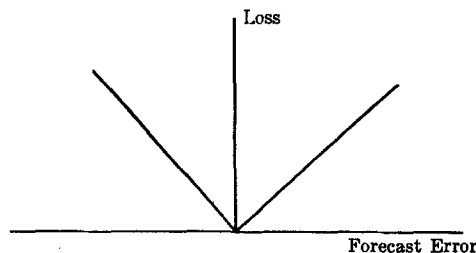

FIGURE 2

Our approach to this problem is, as indicated earlier, pragmatic. We use a model to forecast for time periods for which we have actual observations. The resulting error is observed and this process is repeated until we exhaust the time periods for which data are available. This gives us a time series of observed fore-

casting errors; we calculate and present for each model the mean absolute value and the mean square value of these errors. The model which produces the minimum mean absolute value is regarded as "best" in this paper. The loss function which is implicit here is piecewise linear and is symmetric about the origin (see Fig. 2).

CASE STUDIES IN FORECASTING WITH INTRINSIC MODELS

Curve Fitting

It is an unfortunate and often overlooked fact that a mathematical function which fits a set of data best is not necessarily a good model for forecasting. As an extreme example, it is possible to fit any time series exactly, no matter how complex, with a polynomial of sufficiently high degree. If such a polynomial is used for forecasting, absurd predictions can result. Polynomials are poorly suited for forecasting and their deficiencies are discussed by Cowden in [8].

Exponentially Weighted Moving Average Methods

One important deficiency in ordinary curve fitting is that all observations are given equal weight in determining the parameters of the fitted function. Exponential smoothing methods give more weight to the observations in the immediate past than to those in the more distant past.[4] Specifically, weights which decline in accordance with a geometric series are assigned to the successively more remote observations. This weighting of the observations corresponds to an intuition that recent events contain more important information for forecasting than events that occurred in the more distant past. Still another advantage of these methods is that they can be represented by simple recursive relations. This makes the actual calculation of forecasts extremely easy. A third advantage of exponential smoothing methods is their "robustness." For a certain class of time series, it can be shown that exponential smoothing gives the minimum mean square error forecast of all possible methods utilizing linear combinations of past realizations (see Cogger [7], p. 97). By robustness we mean that when forecasting a time series which does not have the statistical properties of the class referenced above, exponential smoothing often continues to do reasonably well for short-term forecasting. This robustness has been observed for many years and has recently received theoretical support (Cogger [7]).

Simple Exponential Smoothing

The most elementary case of exponential smoothing (the "simple" case) can be represented by the recursive relation

$$\hat{X}_{t+1} = \alpha X_t + (1 - \alpha)\hat{X}_t \tag{1}$$

where X_t is the value of the time series at time t, $\hat{X}_t$ is the forecast value for time

---

[4] Exponential smoothing methods are synonymous with exponentially weighted moving average methods.

t, and α is a parameter to be chosen whose value is between 0 and 1. Although (1) seems to suggest that $\hat{X}_{t+1}$ depends on only the two values X_t and $\hat{X}_t$, it can be shown by expanding (1) that

$$\hat{X}_{t+1} = \alpha \sum_{k=0}^{t-1} (1 - \alpha)^k X_{t-k} + (1 - \alpha)^t X_0, \qquad (2)$$

where X_0 is a starting value. Thus the forecast value $\hat{X}_{t+1}$ is a function of all previous observations in the time series.

An examination of the role of α in equations (1) and (2) indicates that if α is close to 1, most of the weight is assigned to the more recent values of the time series and when α is close to 0 the weight is spread more uniformly over the observations.

A starting value for the calculation made by the recursive relation (1) is sometimes arbitrarily chosen to be the first observation in the time series and sometimes it is taken to be the mean of a small number of the earlier observations. Forecasts are sensitive to the choice of the starting value when the total number of observations in a series is not large. An extensive analysis of this problem is found in Cogger [7] and an empirical study of it appears in Wade [36].

Simple exponential smoothing has been shown to be a useful procedure in forecasting some time series when a large number of short-run forecasts must be produced routinely, and the relatively low cost has been an important aspect of its appeal in a variety of industrial situations. However, simple exponential smoothing did not turn out to be satisfactory in forecasting the time series of residence main stations.

Adaptive Exponential Smoothing

We make a distinction between weakly adaptive and strongly adaptive exponential smoothing models. All exponential smoothing models are adaptive in the sense that forecast values are influenced by the new observations that are included in the time series under study—the forecast values, in short, show some adaptation to this new information. However, in the models discussed in the preceding section the smoothing parameter α is a constant. We call this a weakly adaptive model. In contrast to this, a strongly adaptive model is one in which the smoothing parameter is assigned a value dynamically by making it a function of the sequentially observed forecasting error. In these models, there is adaptation not only to the new information but also through the change in the value of the smoothing parameter itself. The latter type of model we will also call adaptive exponential smoothing; the model now reported on here is discussed in Dunn, Williams, and Spivey [12] and was originally developed by Trigg and Leach [32]. The smoothing constant α of equation (1) is replaced by the function $\alpha(t)$, where

$$\alpha(t) = \left| \frac{SE(t)}{SAE(t)} \right| \qquad (3)$$

$$\text{Smoothed Error} = SE(t) = \gamma e(t) + (1 - \gamma)SE(t - 1) \qquad (4)$$

$$\text{Smoothed Absolute Error} = SAE(t) = \gamma|e(t)| + (1 - \gamma)SAE(t - 1)$$

$$e(t) = X_t - \hat{X}_t \qquad (5)$$

where γ, called the adaptive smoothing constant, is in the interval $0 < \gamma < 1$ and is to be chosen.

Given this function $\alpha(t)$, we then have the following model, which is analogous to simple exponential smoothing,

$$\hat{X}_{t+1} = \alpha(t)X_t + (1 - \alpha(t))\hat{X}_t . \qquad (6)$$

When several large errors of the same algebraic sign occur, it is clear that SE and SAE will give a value for $\alpha(t)$ near 1. This results in a heavy weighting of the recent observations. On the other hand, when forecast errors tend to alternate signs and to be of approximately the same magnitude, SE tends to zero, SAE tends to some nonzero value, and $\alpha(t)$ approaches zero. This gives a model which weights recent observations less heavily and assigns instead more weight to the past history.[5]

The choice of the best value of γ raises the following question: should one choose a large value of γ so as to produce a model that adapts quickly to secular changes in a time series or a smaller value of γ which yields more stable forecast values? There is no generally satisfactory answer to this question; however, since forecasts can be generated quickly with this model, one can experiment with a variety of values of γ, examine the forecast errors that result, and choose a value for γ which minimizes the forecast error criteria indicated earlier, at least among the values of γ that one has examined. Again, it is clear that a knowledge of the nature of the time series one is attempting to forecast is essential.

The use of adaptive exponential smoothing on the residence main station data, although superior to simple exponential smoothing, did not produce forecasts that were sufficiently good.[6] A more extensive analysis of the data was then made.

Data Analysis and Refined Forecasts

An investigation of the main station data indicated that it is in fact the sum of two other time series: the monthly series of telephone connects and the monthly series of disconnects. More precisely, if $MS(t)$ denotes main stations at time t,

---

[5] For a more detailed discussion of adaptive exponential smoothing models see Dunn, Williams, and Spivey [12].

[6] A variant of the Trigg and Leach model, adaptive exponential smoothing with lag, has been indicated by Shone [29]. In this approach

$$\alpha(t) = \frac{SE(t - 1)}{SAE(t - 1)},$$

instead of (3). This guards against the undesirable influence generated by occasional sharp "spikes" in the data and may be useful for some problems. This variant was tried for the telephone data and the amount of improvement it offered was negligible.

Con(t) and Dis(t) denote, respectively, the connects and disconnects at time t, then

$$MS(t) = MS(t - 1) + Con(t) - Dis(t) .\tag{8}$$

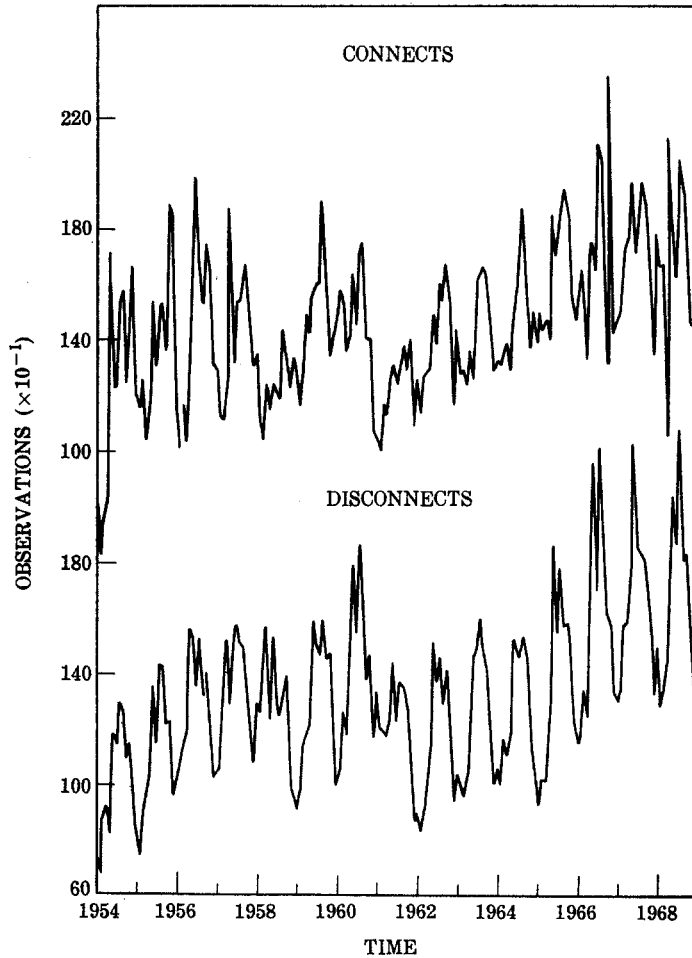The time series of connects and disconnects appear in Figure 3.

FIGURE 3.   RESIDENCE CONNECTS AND DISCONNECTS FOR FLINT

Both of these time series displayed strong seasonal patterns but the patterns differed from each other.   The periodic behavior of these time series was further confirmed by an examination of their sample spectrum and autocorrelation functions.   Roughly speaking, the spectrum of a time series can be regarded as displaying the amount of variance in the series associated with various periodicities.   For purposes of illustration, the sample spectrum of monthly changes (first differences) in Flint main stations is shown in Figure 4.   The high power near a frequency of

84 cycles per month corresponds to the twelve-month periodicity. The large ordinate at a frequency of zero reflects the trend in the series.

Analyzing the spectrum provides one with an extremely sensitive means of detecting periodic behavior in a stationary time series providing the history of the time series is sufficiently long.[7] Unfortunately, most time series in economics are neither stationary nor long so spectral techniques have not proved to be as useful in analyzing economic time series as has been the case in some of the engineering fields.[8]

Applying adaptive exponential smoothing to the series of connects and disconnects and combining the results to obtain a forecast of residence main stations produced better results than did the use of this model on the latter series, but the forecasts still were not accurate enough.

Since the time series had strong seasonal patterns and since no time-series methods appear to accommodate seasonal influences well, a further data decomposition was experimented with. The series of connects was decomposed into twelve separate monthly series in accordance with the conjecture that for a strongly seasonal series, a January observation is perhaps more similar to January observations of other years than to observations for the adjacent months in time of December and February. The series of disconnects was decomposed in the same way, and each of the twenty-four monthly time series was forecast using adaptive exponential smoothing. It should be noted that each series has its own adaptive smoothing function $\alpha(t)$ and its own model. These forecasts were then aggregated into a forecast for residence main stations, producing a forecast with an appreciably lower average absolute forecast error than any of the preceding methods. See Figure 5.

A Further Extension: The Box-Jenkins Methodology

Most of the procedures described above had their origin in attempts to deal with the practical problem of generating better forecasts. These procedures were

---

[7] The probability structure of the time series
$$\xi(t) = \{\ldots, X_{-1}, X_0, X_1, X_2, \ldots\}$$
is considered specified by the set of all finite dimensional distribution functions
$$F_{t_1, t_2, \ldots, t_n}(X_1, X_2, \ldots, X_n) = P\{\xi(t_1) < X_1, \xi(t_2) < X_2, \ldots \xi(t_n) < X_n\},$$
where $t_1, t_2, \ldots, t_n$ are any n elements of the index set $\{\ldots -1, 0, 1, 2, \ldots\}$. A time series is said to be stationary if
$$F_{t_1, t_2, \ldots, t_n}(X_1, \ldots, X_n) = F_{t_1+\tau, t_2+\tau, \ldots, t_n+\tau}(X_1, X_2, \ldots, X_n)$$
where τ is any integer. In particular, this definition means that all one-dimensional distribution functions $F_t(X)$ do not depend on t, so that the time series must have a constant mean value. All two-dimensional distributions must depend only on the difference $t_2 - t_1$, so that the autocorrelation function $\rho(\tau)$ must be a function of τ alone and not of t.

This is a strong condition and is rarely satisfied, even approximately, by economic time series.

[8] Difficulties in estimating the spectral density function have also been a major barrier to the application of important theoretical results of Kolmogoroff [22] and Weiner [38] in the forecasting of time series generally.
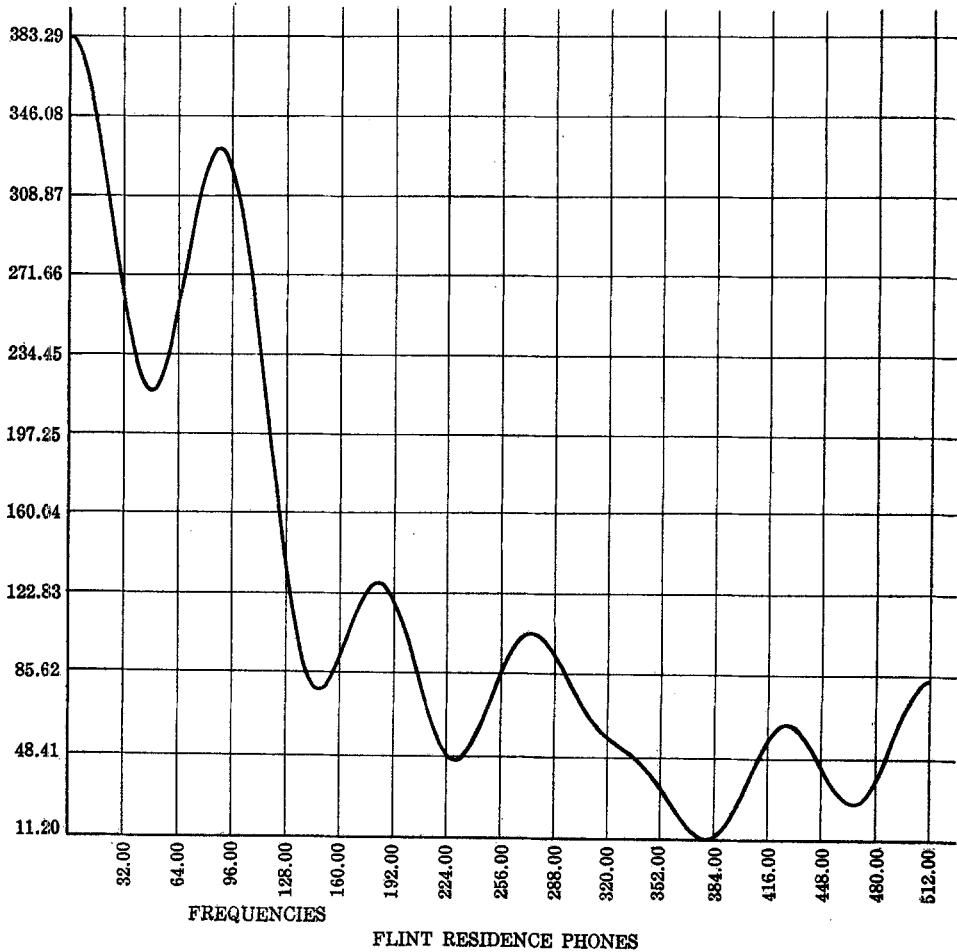
FLINT RESIDENCE PHONES

FIGURE 4. POWER SPECTRA OF FIRST DIFFERENCE

seen originally to have worked well on some problems; they were later refined and extended, mathematical generalizations were developed, and statistical properties were studied. The development of the Box-Jenkins methodology follows a path which is the reverse of this: building on the early theoretical work of Yule [39], Walker [37], and others, Box and Jenkins [3] were able to develop a procedure for forecasting whose statistical properties were worked out in advance. Thus the forecaster can make use of a procedure which utilizes theoretical results which had not previously been exploited.

In a strict sense, the Box-Jenkins procedures do not produce a forecasting model but rather a methodology for constructing a model, estimating its parameters from data, and analyzing its forecasting accuracy.

The theory underlying the probabilistic analysis of time series draws heavily
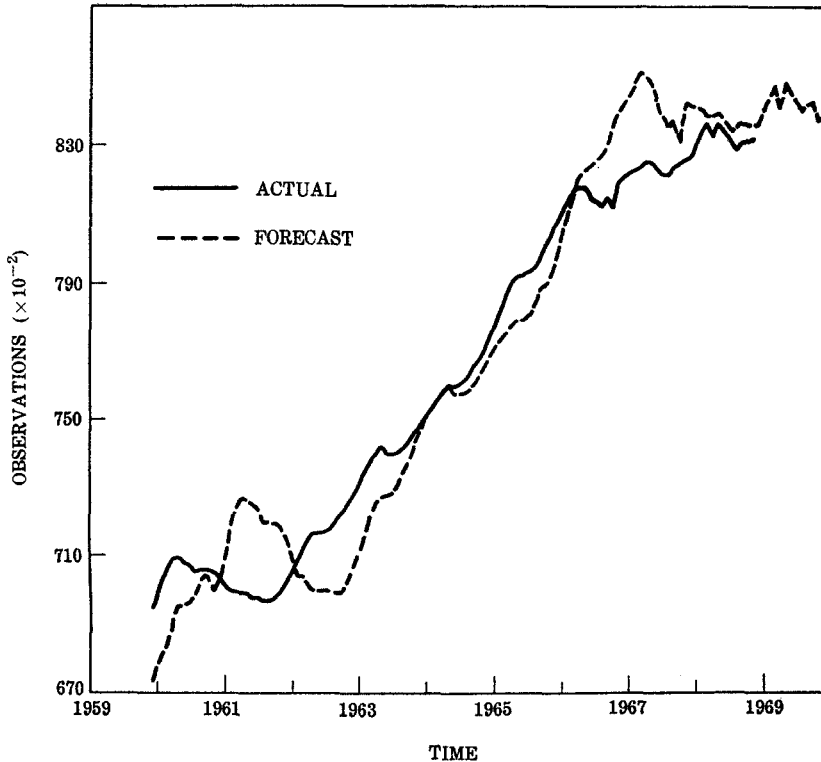
FIGURE 5. FORECAST OF MAIN STATIONS FOR FLINT

on the concept of the autocorrelation function.[9]   In particular, for a large class of stationary time series an optimal linear predictive relation is completely defined by the autocorrelation function of the series.   A basic feature of the Box-Jenkins procedure is an examination of the sample estimator of this function.   By regarding the time series as an autoregressive process with a moving average residual, one can show that there is a functional relation between the value of the autocorrelation function with lag k and the values of the parameters of the forecasting model.[10] The procedure is not clear cut because most economic time series one encounters are not in fact stationary and because the sampling properties of the autocorrelation

---

[9] The autocorrelation function with lag k is given by

$$\rho(k) = \frac{E[(X_t - \mu)(X_{t-k} - \mu)]}{\sqrt{E[(X_t - \mu)^2]E[(X_{t-k} - \mu)^2]}} = \frac{\text{Cov}\,[X_t, X_{t-k}]}{\sigma_x^2}.$$

[10] An autogressive process of order p with a moving average residual of order q is defined as

$$X_t = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + \varepsilon,$$

where the random error term ε is given by

$$\varepsilon = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q}$$

where the ε_j are mutually uncorrelated random variables from a fixed distribution with a mean of zero and finite variance.

function are complex. Box and Jenkins suggest a four-stage, iterative forecasting procedure to deal with these difficulties.

Stage 1. A useful class of models is postulated, for example, the autoregressive-moving average class.

Stage 2. By examining the sample autocorrelation function or the sample estimate of the spectrum, using knowledge of the data-generating process, and so forth, a particular model is chosen from the class of stage 1.

Stage 3. The tentative model is fitted to the data and its parameters estimated.

Stage 4. Diagnostic checks are made on the residuals from the model with the object of discovering systematic lack of fit and identifying the causes. If such an inadequacy is discovered, a modification is made to the tentative model and the procedure is repeated starting with step 2.

The Box-Jenkins procedure also employs various transforms and filters in order to force the time series into a form consistent with the theoretical notion of a stationary autoregressive-moving average random process.

This procedure was applied to the time series of connects and disconnects. We describe how this was done by discussing the situation first for the former and then considering the latter.

Examination of the sample autocorrelation function of the first and higher order differences of the time series of connects suggested that one could regard the time series of first differences as approximately stationary.[11] That is, the series $\{d_t\}$, where

$$d_t = \text{Con}_t - \text{Con}_{t-1},$$

appeared to have autocorrelation properties similar to an autoregressive moving average random process which is known to be stationary. The sample estimate of the autocorrelation function of d_t, together with the estimated standard errors of the estimates, are shown in Table 1.[12] Note that with the exceptions of lag 1 and lag 12, all estimates of the autocorrelation function are less than two standard deviations from 0. These autocorrelations may be accounted for by assuming that

---

[11] The sample estimate $\hat{\rho}_k$ of the autocorrelation function ρ_k at lag k is given by

$$\hat{\rho}_k = \frac{\hat{r}_k}{\hat{r}_0},$$

where $\hat{r}_k = (1/N) \sum_{t=1}^{N-k} (X_t - \bar{X})(X_{t+k} - \bar{X})$ and $\bar{X}$ is the mean of the time series. The actual calculations were performed on the logarithms of the time series of connects (no monthly decomposition was used with the Box-Jenkins procedure).

[12] The standard error of $\hat{\rho}_k$, the value of the estimate of the autocorrelation function at lag k, is obtained from the following approximate relationship,

$$\text{Var}(\hat{\rho}_k) = \frac{1}{N}\{1 + 2\sum_{i=1}^{k} (\hat{\rho}_i)^2\}.$$

This result, due to Bartlett, is cited in Box and Jenkins [3], p. 34.

TABLE 1

ESTIMATES OF AUTOCORRELATION FUNCTION OF

FIRST DIFFERENCES OF CONNECT TIME SERIES

Lag in Months													Estimated Standard Error
1–12:	−.37	−.03	.05	−.04	−.14	.03	−.04	−.09	.04	−.07	.09	.22	.11
13–24:	−.09	.12	−.05	−.08	−.05	.05	−.12	−.04	−.05	−.06	.22	.05	.12
25–36:	−.07	.23	−.17	−.04	.05	−.15	−.02	−.03	−.03	.07	.07	.08	.12
37–48:	.10	.00	−.15	.12	−.09	−.07	−.03	−.04	−.07	.17	−.06	.05	.13

$$d_t = \varepsilon_t - \lambda_1 \varepsilon_{t-1} - \lambda_{12} \varepsilon_{t-12} ,$$

where ε_t is defined in footnote 10.

Thus the forecasting equation is

$$\hat{X}_{t+1} = X_t + \varepsilon_{t+1} - \lambda_1 \varepsilon_t - \lambda_{12} \varepsilon_{t-11} ,$$

where the value of ε_{t+1} is set equal to its expected value of 0 and the parameters λ_1 and λ_{12} are yet to be determined.

To develop forecasts for lead times greater than one period in advance, the forecasting equation is used recursively. Unknown values of X_t are replaced by $\hat{X}_t$ and unknown values of ε_t are replaced by 0. A modification was made in the usual Box-Jenkins method of estimating the parameters λ_1 and λ_{12}. In order to be consistent with the forecasting results of other models discussed in this paper, λ_1 and λ_{12} were chosen so as to minimize the mean absolute value of the sequentially observed "twelve step ahead" forecast error of the time series of main stations. The usual procedure, incidentally, is to select the parameters subject to minimizing the "one step ahead" mean-squared forecast error criterion. The latter, together with a normality assumption concerning the errors, enables one to state that the resulting estimates are asymptotically maximum likelihood estimates of the parameters.

The forecast equation resulting from the minimization of the absolute value of forecast error is

$$\hat{X}_{t+1} = X_t - .30\varepsilon_t - .0\varepsilon_{t-11} .$$

An examination of the time series of forecast errors generated by this method indicated that they have the same statistical properties as uncorrelated random deviates. As a result, the model was regarded as satisfactory.

A similar analysis applied to the time series of disconnects produced the forecasting equation

$$\hat{X}_{t+1} = X_t - .20\varepsilon_t + .70\varepsilon_{t-11} - .20\varepsilon_{t-12} .$$

Combining the forecasts of the connects and disconnects to develop a forecast for main stations produced the forecasts shown in Figure 6. The mean absolute error of this model was 1344.0.

Hybrid Models

We indicated earlier that the combining of an intrinsic and an extrinsic forecasting strategy in a hybrid model has some appeal and that surprisingly little formal work has been done in this area. Two possibilities suggest themselves immediately. One could retain an essentially extrinsic forecasting strategy and take advantage of an intrinsic model in some way. For example, in an econometric model one might choose to forecast an exogenous variable by time-series methods. Alternatively, one could use an intrinsic model and incorporate information on one or more
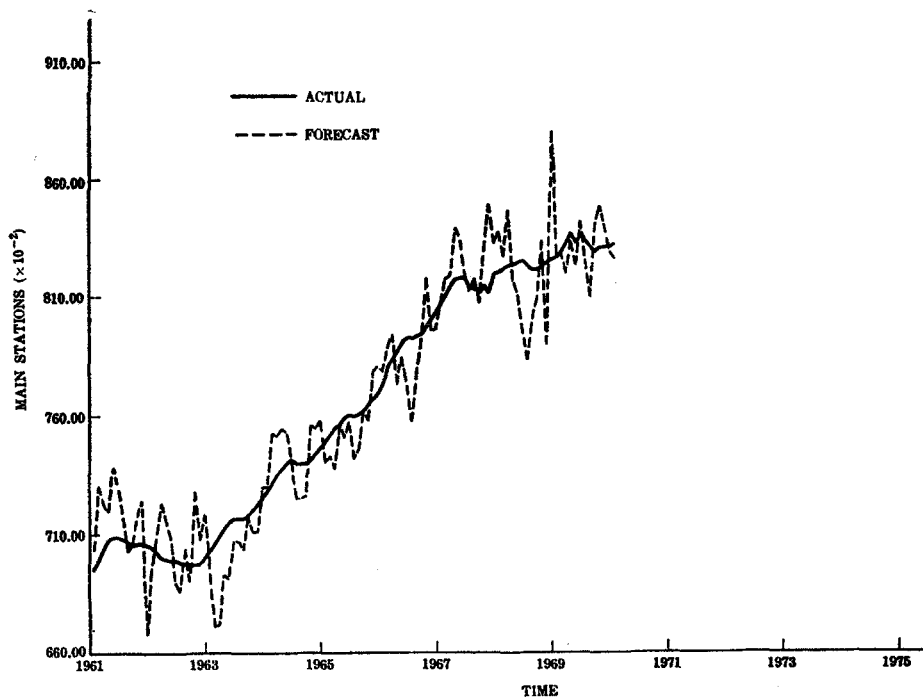
FIGURE 6. FORECASTS OF FLINT MAIN STATIONS
USING BOX-JENKINS

related variables into it.

One should, moreover, be able to improve the performance of a forecasting model if an extrinsic influence can be found which leads the variable one is trying to forecast. In the spirit of a case study, we would like to present briefly an account of a crude attempt that was made by one of the authors in association with others (see Dunn, Williams, and Spivey [12]) that seems to confirm that using an extrinsic or exogenous variable in what is otherwise an intrinsic model holds promise for future development.

Returning to the problem of forecasting telephone main stations, it is an obvious conjecture that new household formation is a leading indicator. If the sequence $\{Y_t\}$ represents observations on new household formation in Flint, then one way of incorporating the exogenous variable is to begin with the adaptive exponential smoothing relation

$$\hat{X}_t = \alpha(t)X_{t-1} + [1 - \alpha(t)]\hat{X}_{t-1}$$

and modify it into the relation

$$\hat{X}_t = \alpha(t) X_{t-1} + [1 - \alpha(t)]\hat{X}_{t-1} + \beta\alpha(t)Y_{t-1} \qquad (9)$$

where $\alpha(t)$ is the smoothing function defined in (3) and where β is a constant (not

restricted to be between 0 and 1) chosen according to the criterion of minimizing the mean absolute deviation of forecast errors. By adaptively assigning the value of $\alpha(t)$ one can include more of the effect of the observations Y_i on the exogenous variable when forecast errors are large and reduce its contribution during relatively stable intervals.
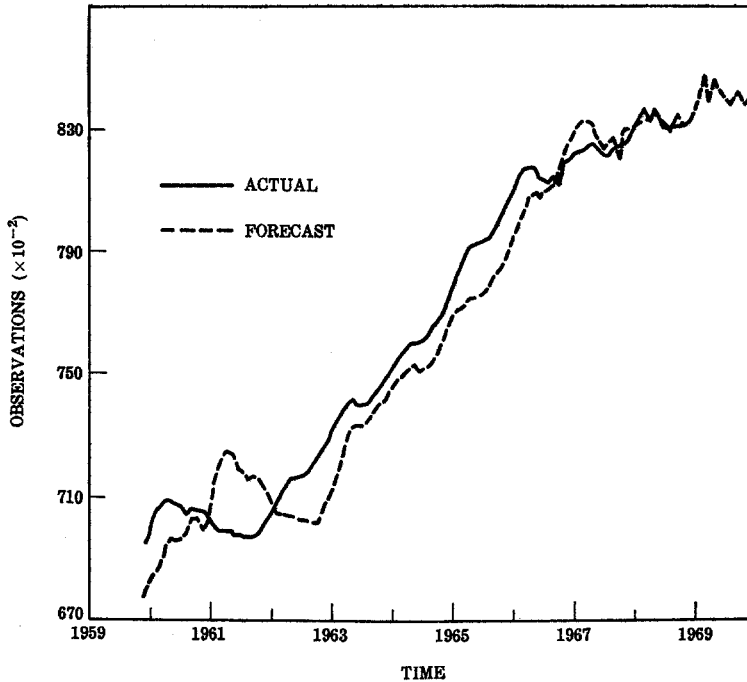

FIGURE 7. FORECAST OF MAIN STATIONS FOR FLINT
USING EXOGENOUS DATA

Using equation (9) on the decomposed monthly series and aggregating back to a forecast for residence main stations resulted in forecasts for which the forecast error was 14 per cent less than for that for the best of the other adaptive forecasting models. This is all the more surprising when it is observed that a monthly time series on new household formation is not available for Flint and that observations taken at annual intervals were arbitrarily allocated to months assuming a constant growth rate over the months of a year and the resulting series used as the exogenous variable. An illustration of forecasts with this model and the time series of main stations is shown in Figure 7.

The equation (9), of course, represents only a crude assimilation of exogenous information. One is naturally led to ask if more sophisticated procedures can be used and if their theoretical properties can be developed. Research into these questions by the authors is presently underway.

Summary

In Table 2 the mean absolute error for those models which performed best in the context of the telephone demand problem are shown, together with corresponding parameter values. One sees that the hybrid model produced the most accurate forecasts. It is interesting to observe that this model was superior to the vastly more complex Box-Jenkins procedure.

TABLE 2

SUMMARY OF FORECASTS

Forecasting Technique	Mean Absolute Error (12-month forecast)	Parameter Values
Trigg and Leach adaptive exponential smoothing with decomposition	1226	$\gamma = .9$
Decomposition with Shone variant of Trigg and Leach adaptive exponential smoothing	1161	$\gamma = .9$
Hybrid model using exogenous variable	986	$\gamma = .1$ $\beta = 10.74$
Box-Jenkins procedure	1344	Connects: $\lambda_1 = .30$, $\lambda_{12} = 0$ Disconnects: $\lambda_1 = .20$, $\lambda_{12} = -.70$ $\lambda_{13} = -.20$

We conclude by stating our belief that intrinsic models, together with supporting data analyses, have much to offer the regional scientist. Although it cannot be said that these models will be superior in a large number of cases to other models currently available, their relatively modest data requirements and ease of implementation provide a useful addition to the largely associative models that appear to be widely used today.

REFERENCES

[1] Anderson, R. L., Jr. "A Note on Economic Base Studies and Regional Econometric Forecasting Models," *Journal of Regional Science*, X, No. 3, 1970, pp. 325–333.

[2] Bell, F. W. "An Econometric Forecasting Model for a Region," *Journal of Regional Science*, VII, 1967, pp. 109–127.

[3] Box, G. E. P., and Jenkins, G. M. *Time Series Analysis*. San Francisco, Calif.: Holden-Day, Inc., 1970.

[4] Brown, R. G. *Smoothing, Forecasting and Prediction*. Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1965.

[5] Chandmal, A., and Jayaraj, C. "A Communication on Adaptive Smoothing Using Evolutionary Spectra," *Management Science*, XVIII, No. 1, 1971, pp. 112–113.

[6] Chow, W. M. "Adaptive Control of the Exponential Smoothing Constant," *Journal of Industrial Engineering*, XVI, No. 5, 1965, pp. 314–17.

[7] Cogger, Kenneth. "Statistical Foundations of General Order Exponential Smoothing with Implications for Applied Time Series Analysis," unpublished Ph.D. dissertation, The University of Michigan, 1971, 270p.

[8] Cowden, D. J. "The Perils of Polynomials," *Management Science*, IX, No. 4, 1963, pp. 542–550.

[9] Draper, N. R., and Smith, H. *Applied Regression Analysis*. New York: John Wiley and Sons, 1966.

[10] Dunn, D. M. "Local Area Forecasting: An Adaptive Approach," unpublished Ph.D. dissertation, The University of Michigan, 1970, 343p.

[11] Dunn, D. M., W. H. Williams, and W. A. Spivey. "Forecasting the Local Area Demand for Telephones," *Proceedings of the American Statistical Association*, 1970, pp. 470–474.

[12] ———. "Analysis and Prediction of Telephone Demand in Local Geographical Areas," *The Bell Journal of Economics and Management Science*, II, No. 3, 1971, pp. 561–576.

[13] Glickman, N. J. "An Econometric Forecasting Model for the Philadelphia Region," *Journal of Regional Science*, XI, No. 1, 1971, pp. 15–32.

[14] Goodman, M. L., and W. H. Williams. "A Simple Method for the Construction of Empirical Confidence Limits for Economic Forecasts," *Journal of the American Statistical Association*, December, 1971.

[15] Granger, C. W. J., and M. Hatanaka. *Spectral Analysis of Economic Time Series*. Princeton: Princeton University Press, 1957.

[16] Hannan, E. J. *Time Series Analysis*. London: Methuen & Co., 1960.

[17] Hooper, J. W., and A. Zellner. "The Error of Forecast for Multivariate Regression Models," *Econometrica*, XXIX, No. 4, 1961, p. 12.

[18] Isard, W. *et al*. *Methods of Regional Analysis: An Introduction to Regional Science*. Cambridge, Mass.: The M.I.T. Press, 1960.

[19] Jenkins, G. M., and D. C. Watts. *Spectral Analysis*. San Francisco, Calif.: Holden-Day, Inc., 1968.

[20] Kendall, M. G. "A Theorem in Trend Analysis," *Biometrica*, XLVIII, No. 4, 1961, p. 224.

[21] Klein, L. "Whither Econometrics," *Journal of the American Statistical Association*, LXVI, No. 334, 1971, pp. 415–421.

[22] Kolmogoroff, A. N. "Interpolation and Extrapolation of Stationary Random Sequences," *Izv. Akad. Nauk. SSSR*, Ser. Mat. 5, 3, 1941.

[23] Malinvaud, E. *Statistical Methods of Econometrics*. 2nd rev. ed. Amsterdam: North Holland Publishing Co., 1970.

[24] Mattila, John. *Estimating Metropolitan Income, Detroit 1950–1969*. Detroit, Mich.: Wayne State University, Center for Urban Studies, January, 1970.

[25] Meyer, R. F. "An Adaptive Method for Routine Short-Term Forecasting," *Proceedings of the Third International Conference on Operations Research*, Oslo, Norway, 1963, pp. 882–893.

[26] Miernyk, W. *The Elements of Input-Output Analysis*. New York: Random House, 1965.

[27] Mincer, J., ed. *Economic Forecasts and Expectations*. National Bureau of Economic Research, 1969.

[28] Rao, A. G., and A. Shapiro. "Adaptive Smoothing Using Evolutionary Spectra," *Management Science*, XVII, No. 3, 1970, pp. 208–218.

[29] Shone, M. L. "Exponential Smoothing with an Adaptive Response Rate," *Operational Research Quarterly*, XVIII, No. 3, 1967, pp. 318–319.

[30] Theil, H., and S. Wage. "Some Observations on Adaptive Forecasting," *Management Science*, X, No. 2, 1964, pp. 198–206.

[31] Tiebout, C. M. *The Community Economic Base Study*. Supplementary Paper No. 16, Committee for Economic Development, 1962.

[32] Trigg, D. W., and A. G. Leach. "Exponential Smoothing with an Adaptive Response Rate," *Operational Research Quarterly*, XVIII, No. 1, 1967, pp. 53–59.

[33] Tukey, J. W. "Discussion Emphasizing the Connection between Analysis of Variance and Spectral Analysis," *Technometrics*, III, No. 1, 1961, pp. 191–219.

[34] ———. "The Future of Data Analysis," *Annals of Mathematical Statistics*, XXXIII, 1962, pp. 1–62.

[35] Tukey, J. W., and M. B. Wilk. "Data Analysis and Statistics: An Expository Overview," *Proceedings of the Fall Joint Computer Conference*, 1966, pp. 695–709.

[36] Wade, R. C. "Techniques for Initializing Exponential Smoothing Forecasts," *Management Science*, XIII, 1967, pp. 601–602.

[37] Walker, G. "On Periodicity in a Series of Related Terms," *Proceedings of the Royal Society*, A131, 1931, p. 518.

[38] Weiner, N. *Extrapolation, Interpolation and Smoothing of Stationary Time Series*. New York: John Wiley and Sons, 1949.

[39] Yule, G. U. "On a Method of Investigating Periodicities in a Disturbed Series, with Special Reference to Wolfer's Sunspot Numbers," *Phil. Trans.*, A226, 1927, p. 267.