AT&T Bell Laboratories Technical Journal
Vol. 63, No. 5, May–June 1984
Printed in U.S.A.

On the Accuracy of Forecasting Telephone Usage Demand

By M. N. YOUSSEF*

(Manuscript received March 3, 1981)

This paper examines the accuracy of telephone-usage-demand forecasts produced by three state-of-the-art forecasting techniques: Box-Jenkins, Akaike state space, and an autoregressive spectrum estimation. The study considers 35 actual monthly demand time series and 300 simulated realizations. Principal results are that: (1) correct identification of the nonstationary behavior of telephone demand is crucial to forecast performance, (2) overparameterization or underparameterization of the stationary aspects of a process has little or no impact on the accuracy of the forecast, and (3) forecasts based on the naive random-walk method compare favorably to those produced by sophisticated techniques. Also, strengths and weaknesses of the investigated techniques are revealed through data analysis. It is further argued that the traditional method for assessing the accuracy of the usage forecast based on average busy season quantities is biased towards underforecasting.

I. INTRODUCTION

Errors in forecasting per-telephone usage demand at the switching-machine level can be quite costly for the Bell operating companies. It has been estimated that a net 3-percent overforecast would yield an increase in capital requirements of about $800 million.[1] The result of underforecasting is degraded service.

The usage variable to be projected is known as the hundred call seconds per main station (CCS/M).† This variable is required to

* AT&T Bell Laboratories.

†One CCS is one hundred call seconds of usage per hour and is equal to 1/36 Erlang.

determine central office relief timing and sizing strategies. In this regard forecasts are needed up to five years ahead.

Historically, the most commonly used CCS/M forecasting technique in the Bell operating companies has been linear regression trending. The result has often been a large forecast error. In this paper, we examine three state-of-the-art time-series techniques. These are the Autoregressive Integrated Moving Average (ARIMA) class of models and the Box-Jenkins philosophy, a state space method developed by Akaike, and a robust autoregressive spectrum estimation and factorization approach.

The Box-Jenkins technique provides a procedure for model building.[2] But, because the procedure requires user judgment, it can lead different users to identify different models for the same set of data. The question naturally arises as to what situations lead to different identified models and what impact this has on forecasting accuracy.

On the other hand, attempts have been made to eliminate judgment by developing procedures that automatically determine the order of the most representative model. The state space approach presented in this paper and its information criterion (known as the AIC) have been advocated to do just that, assuming that the given process is stationary.[3]

Furthermore, efforts were spent in seeking procedures that may circumvent the identification stage entirely. The nonparametric method for estimating the spectrum, and hence the coefficients of autoregressive models, is a result of these efforts. It is examined here since it may provide a viable alternative.[4]

As a first step in assessing the three techniques, we selected 35 time-series data on CCS/M from more than 100 series that were available. In this selection we avoided time series that were contaminated with many outliers or with large jumps. The selection was strictly based on visual inspection of the time-series graphs.

Limiting the investigation to "clean" data does not hinder the significance of its findings. Our primary objective is to answer the question: Is there a method better than a naive one that can improve the CCS/M forecast even for the most regular offices? Moreover, concerning outliers, steps have been taken to guarantee better data-collection procedures via reliable mechanization. As for jumps, we were able to associate a causative factor with each jump observed in the data. Such information ought to be incorporated in the forecasting mechanism.

Results obtained from applying these techniques to the selected time series show that none of the techniques was able to provide a marked improvement in forecast accuracy—even for 1 year ahead—when compared to a simple-minded forecast that uses data from 1 year past. Moreover, the models identified via the state space approach were different from those identified by the Box-Jenkins technique. Existing literature concerning the performance of the AIC criterion in identifying the order of a pure autoregressive model indicates that the AIC has a tendency to overparameterize.[5] Our results show that the criterion mostly underparameterizes and is inconsistent in general.

As a first impression, one may attribute these findings to many factors. Some possibilities include the short history data, outliers, failure to reduce the data to stationary time series, nonconstant variance of the noise process and/or the applicability of the criterion used in comparing forecasting accuracy of the various methods.

The Bell operating companies measure of forecast accuracy for the CCS/M is the percent difference between a forecast and an actual for a quantity known as Average Busy Season (ABS). The ABS is defined as the average of the highest three observations in a year. Forecast accuracy is only relevant for ABS values. It is the ABS forecast that is used for planning and engineering.

A simulation experiment was conducted to verify and explain results obtained from analyzing the CCS/M data. Three typical processes identified from the CCS/M time series were chosen. In the simulation, one hundred realizations were generated from each of the three processes. Using the simulated data to assess the various approaches, findings similar to those obtained from analyzing actual data were disclosed. These findings, however, provided additional insight into a number of aspects of the problem.

This paper is organized as follows. Section II outlines the three investigated techniques. Section III discusses our experience in applying these techniques to the CCS/M data, and illustrates, where possible, the effect of overfitting, underfitting, and misspecification of model parameters on the forecast accuracy. Section IV describes the simulation experiment and assesses the performance of the three approaches in an artificially structured ideal environment for the CCS/M processes. Section V summarizes the results and outlines a number of observations that were noted during the investigation.

II. METHODOLOGY

The three investigated approaches for the analysis and forecasting of the CCS/M variable are all claimed to be powerful. They are also complicated and expensive (in varying degrees). They depend heavily on elaborate software packages that are not normally available in standard computer systems.

Our reason, then, for investigating these approaches was to understand the properties of the data in order to suggest a simple method. Hypothetically, using state-of-the-art methods would allow for quantifying the maximum forecast accuracy that can be achieved, or the maximum loss that may be incurred due to simplification.

In the next three subsections, we outline these methods. The following presentation assumes that the reader is familiar with autoregressive and moving average processes.

2.1 The Box-Jenkins approach for seasonal data

For seasonal time series of period S Box and Jenkins suggest representation by members of the following class

$$\phi_p(B)\Phi_P(B^S)(1 - B)^d(1 - B^S)^D(z_t - \mu) = \theta_q(B)\Theta_Q(B^S)a_t, \quad (1)$$

where

z_t = the time series data at time t

μ = the location parameter

B = the backshift operator such that $B^k z_t = z_{t-k}$

a_t = independent random variable with mean zero and variance σ_a^2, known as white noise

d and D = the degrees of differencing required for achieving process stationarity

$\phi_p(B)$, $\Phi_P(B)$, $\theta_q(B)$ and $\Theta_Z(B)$ = polynomials in B of order p, P, q, and Q, respectively.

The class of models in (1) is known as multiplicative seasonal Autoregressive Integrated Moving Average (ARIMA) of order $(p, d, q) \times (P, D, Q)_S$.

The Box-Jenkins approach for fitting models of the form (1) involves a three-part cycle of identification, estimation, and diagnostic checking. The cycle is ended once an adequate model is derived. The basic tools for model identification are two descriptive functions known as the sample Autocorrelation Function (ACF) and the sample Partial Autocorrelation Function (PACF). Nonstationarity of a series is identified from the behavior of its ACF. For nonstationary series, the serial correlation in the ACF remains large at large lags. In this situation, the integrated part of the ARIMA representation allows for differencing to induce stationarity. Once stationarity is achieved, the problem is to select reasonable values for p, q, P, and Q. This is also done by examining the shape of the ACF (and PACF) of the stationary series. In the estimation stage, the parameters ϕ's, Φ's, θ's, and Θ's of the identified model are calculated on the basis of the minimization of the sum of squared errors $\{a_t\}$.

After the model is fitted, the residuals $\{\hat{a}_t\}$ are checked for whiteness. If no pattern is detected in the autocorrelations of residuals, the assumption that the a_t's are white noise is accepted. If, on the other hand, a pattern is observed, such a pattern would provide information on how to modify the model. In the latter case, the identification, estimation, and diagnostic checking cycle is repeated.

We note from (1) that the Box-Jenkins approach provides a capability for representing a wide class of nonstationary time series. Through differencing an appropriate number of times, the nonstationary behavior can be removed. The use of differencing, however, requires judgment. Nevertheless, we adopted this technique to transform all 35 CCS/M time-series data before employing any of the other approaches. These other approaches deal only with the representation of stationary processes. But, we tested the impact of adopting this technique on the results via the simulation experiment.

2.2 State space approach

While the state space approach can accommodate multivariate systems, for the purpose of this study, we address the univariate case only. From control theory, a linear, discrete-time, time-invariant system can be represented by

$$v_{t+1} = Fv_t + gu_{t+1}$$
$$x_t = hv_t, \tag{2}$$

where v_t is an $r \times 1$ state vector, and $\mathbf{F}$, $\mathbf{g}$, $\mathbf{h}$ are $r \times r$, $r \times 1$, and $1 \times r$ matrices, respectively. This representation assumes x_t (scalar) to be the output and u_t to be the input of the system.

Akaike showed that, when x_t and u_t are stochastic Gaussian processes an analogous representation to (2) exists.[6,7] The new representation is called Markovian and can be obtained from the analysis of canonical correlations between the set of the present and future output and the set of the present and past input. We consider in this paper the case where there is a feedback from the output to the input, that is, when $u_t = x_t$. This Markovian representation takes the form

$$v_{t+1} = Fv_t + ga_{t+1}$$
$$x_t = hv_t, \tag{3}$$

where a_{t+1} is the innovation of x_t at time $t + 1$. It is defined by

$$a_{t+1} = x_{t+1} - x_{t+1|t}. \tag{4}$$

Here $\{a_t\}$ is white noise with autocovariance $c_\tau = \mathbf{E}(a_t \cdot a_{t-\tau})$. Also, $x_{t+1|t}$ is the projection of x_{t+1} onto the linear space $[R(t-)]$ spanned by the components of the present and the past $(x_t, x_{t-1}, \cdots)$. It is the one-step-ahead predictor of x_t. The space that is spanned by the components of the predictors $x_{t+i|t}$ is called the predictor space $[R(t+|t-)]$. The components of the state vector, v_t, are elements of this space, which is assumed to have a finite dimension. In this regard v_t provides a specification of the predictor space.

Since any basis of the predictor space can play a role of the state vector, **F**, **g**, and **h** in (3) can have different structures. However, a state vector whose elements are the first maximum set of independent components of the predictor space has the smallest possible dimension. This "minimal" representation defines a canonical representation of the system. Its dimension is equal to the number of nonzero canonical correlation coefficients. We are interested in the canonical representation because it corresponds to the concept of parsimony in the ARMA representation, and hence to the problem of identifiability of the process.

To show the relation between the Markovian representation and the ARMA representation, suppose an ARMA model for x_t is given in the form

$$x_t + \phi_1 x_{t-1} + \cdots + \phi_p x_{t-p} = a_t + \theta_1 a_{t-1} + \cdots + \theta_q a_{t-q} \tag{5}$$

or

$$\phi(B)x_t = \theta(B)a_t.$$

The zeros of $\phi(B)$ and $\theta(B)$ are constrained to lie outside the unit circle to ensure stationarity and invertibility of the process. Hence x_t can be expressed as

$$x_t = \sum_{i=0}^{\infty} \psi_i a_{t-i}, \tag{6}$$

where

$$\psi_0 = 1 \qquad \text{and} \qquad \sum_{i=0}^{\infty} \psi_i^2 < \infty.$$

Since $x_{t+i|t} = x_{t+i}$ for $i = 0, -1, -2, \cdots$ and $a_{t+i} = 0$ for $i = 1, 2, \cdots$, eq. (5) can be expressed by

$$x_{t+i|t} + \phi_1 x_{t+i-1|t} + \cdots + \phi_p x_{t+i-p|t} = a_{t+i|t} + \theta_1 a_{t+i-1|t} + \cdots + \theta_q a_{t+i-q|t}. \tag{7}$$

For $i \geqslant q + 1$, the right-hand side of (7) vanishes. Thus for $r \geqslant \max(p, q + 1)$, (7) becomes

$$x_{t+r|t} = -\phi_1 x_{t+r-1|t} - \phi_2 x_{t+r-2|t} - \cdots - \phi_r x_{t|t}, \tag{8}$$

where

$$\phi_i = 0 \qquad \text{for} \qquad r > p.$$

Also, from (6) one can get

$$x_{t+i+1|t+1} - x_{t+i+1|t} = \Psi_i a_{t+1}. \tag{9}$$

From (8) and (9) one can see that the state vector $v_t = (x_{t|t} x_{t+1|t} \cdots x_{t+r-1|t})'$ provides a Markovian representation in the form given by (3), where

$$\mathbf{F} = \begin{bmatrix} 0 & 1 & \cdots & 0 \\ 0 & 0 & & 0 \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ 0 & 0 & & 1 \\ -\phi_r & -\phi_{r-1} & \cdots & -\phi_1 \end{bmatrix}$$

$$\mathbf{g} = [1 \quad \psi_1 \quad \cdots \quad \psi_{r-1}]'$$

$$\mathbf{h} = [1 \quad 0 \quad \cdots \quad 0]'.$$

Here, the ψ's are obtained from the relation

$$\theta_i = \psi_i + \phi_1 \psi_{i-1} + \cdots + \phi_p \psi_{i-p}, \qquad i = 1, 2, \cdots, q,$$

where

$$\psi_j = 0 \qquad \text{for} \qquad j < 0.$$

Thus, given an ARMA representation of a stationary process, one can derive a Markovian representation. Similarly, it can be shown that any stationary process that has a Markovian representation has an ARMA representation. Also, the minimal Markovian representation—known as canonical representation—corresponds to a parsimonious ARMA model.

In practice, the canonical representation needs to be identified from a finite length of a time series. The theoretical canonical correlation coefficients are replaced by the sample canonical correlation coefficients. A criterion is required to determine how small the sample canonical correlation coefficients should be to be judged zero. An information criterion (AIC) was suggested by Akaike in this regard. The AIC criterion is defined by

$$\text{AIC} = -2(\text{maximum log likelihood}) + 2(p + q).$$

An approximation of the likelihood function is calculated from the sample autocovariances. The log likelihood measures the goodness of model fit. The term $2(p + q)$ gives preference to a structure with the least number of free parameters. It was argued that the canonical structure has Minimum AIC (MAICE). There is no theoretical proof of the optimality of the MAICE in this context however.

A procedure for automatically identifying the canonical structure from time-series data is described in Ref. 8. In this procedure, the amount of past to be used in the canonical-correlation analysis is determined from fitting a sequence of autoregressive models. It is equal to the order of the model that has MAICE. The canonical correlations between the past data and the future with increasing number of steps are then calculated. These calculations determine a structure of the state vector by which variables that yield large correlations are included. Usually, the provided structure of the state vector is not necessarily adequate.[7] Besides, estimates provided by this type of analysis are poor. However, this structure and the estimated parameters may be used as initial guesses. Computation of the likelihood function for this structure and all possible choices of the state vector are then performed. The MAICE structure is hence selected.

For maximum likelihood estimation, an approximation was developed since direct maximization of the likelihood function for each possible structure is a formidable task. The developed algorithm is based on Davidon's variance algorithm,[7] which requires the evaluation of the gradient and the inversion of the approximate Hessian of the log likelihood. Our experience with the application of this approach to telephone usage data will be further discussed in Section 3.3.

2.3 Autoregressive spectrum estimation approach

This approach involves the fitting of an autoregressive process of order p—where p is unknown—with the coefficients being estimated in the frequency domain. The power spectral density function and its factorization are used to estimate model parameters. The order, p, may be determined using a stopping rule criterion, such as Parzen's[9] or Akaike's.[10] The following is a brief account of an autoregressive spectral estimation approach. The reader is also referred to the work in Refs. 9, 11, 12, and 13.

Let x_t be a zero mean, discrete time, stationary stochastic process with autocovariance C_τ at lag τ. Then

$$C_\tau = \mathrm{cov}(x_t, x_{t-\tau}). \tag{10}$$

The power spectral density function $f(\omega)$ at frequency ω is defined as

$$f(\omega) = \frac{1}{2\pi} \sum_{\tau} C_\tau e^{-i\omega\tau}, \qquad |\omega| \leq \pi. \tag{11}$$

Since $\{x_t\}$ can be represented by an autoregressive process in the form

$$\sum_{j=0}^{\infty} \phi_j x_{t-j} = a_t, \tag{12}$$

where

$$\phi_0 = 1 \qquad \text{and} \qquad a_t \sim N(0, \sigma_a^2),$$

the spectral density of the process is

$$f(\omega) = \frac{\sigma_a^2}{2\pi} \left| \sum_{j=0}^{\infty} \phi_j e^{ij\omega} \right|^{-2} = \frac{\sigma_a^2}{2\pi[A(\omega)]^2}, \tag{13}$$

where

$$A(\omega) = \sum_{j=0}^{\infty} \phi_j e^{ij\omega}$$

is the transfer function of ϕ_j's.

For a known $f(\omega)$ and an infinite series x_t, factorization of the spectrum provides values for the parameters ϕ_j's. A procedure for factorization of $f(\omega)$ that can determine model coefficients is described in Ref. 4.

In practice, however, the spectral density function is unknown and there is only a finite realization, say of length T, of a time series. The concept of a "windowed" estimate for $f(\omega)$ is used for carrying out the factorization and subsequently the estimation of ϕ_j's. The following discussion outlines the method. Briefly,

$$C_\tau^T = \frac{1}{T} \sum_{t=0}^{T-r-1} x_t x_{t+r} \qquad 0 \leq \tau < T \tag{14}$$

gives the sample covariance function. An estimation of the spectral function can be expressed by

$$f^T(\omega) = \sum_{\tau} \frac{1}{2\pi} C_\tau^T \, k\left(\frac{\tau}{p}\right) e^{-i\tau\omega}, \tag{15}$$

where p is the truncation point that gives the order of the model, and $k(v)$ is called the lag window. Usually, $f^T(\omega)$ and subsequently ϕ_j's are estimated at N points equally spaced in $(0, \pi)$. The equation $f^T(\omega_i)$, $\omega_i = 2\pi i/2N$ for $(i = 0, \cdots, N-1)$ is called the "windowed" estimate of $f(\omega)$.

The spectral analysis and estimation of model parameters are done by three applications of the Fast Fourier Transform (FFT) algorithm.[13] The software for carrying out this analysis is currently available in the Statistical Computing Library (STATLIB).[14]

Our use of the analysis in the frequency domain is limited to the estimation of the model parameters $(\phi_j, j = 1, \cdots, p)$. The advantage of this approach is that the problem of model identification can be avoided for the stationary part of the process.

III. APPLICATION OF METHODOLOGY TO THE CCS/M DATA

Central office CCS/M monthly data were sampled for 35 switching entities. Preliminary analysis of these data was performed to check if the requirements on the methodology used were met. Data transformations were employed where needed. Forecasting models from each of the three techniques were fitted to the transformed data for the period between May 1969 and April 1975. Forecasts were then generated for the next 24 months and converted to ABS values. The percent forecast error—which is defined as the percent difference between actual and forecast for ABS values—for 1 and 2 years ahead were hence calculated. Percent forecast errors were also calculated for the random-walk method, which considers the future ABS of a given series to be its most recent ABS value. This naive method is selected to serve as a benchmark in assessing forecast performance for the other compared methods.

3.1 Preliminary data analysis

As a first step in the analysis, we plotted the raw data for each of the time series. Figures 1 through 3 show time-series graphs for three of the series (A, B, C). We consider the behavior exhibited in Figs. 1 and 2 to be typical of the CCS/M data and that of Fig. 3 to be less typical. The sample autocorrelation function for Series A is also shown in Fig. 4. Preliminary assessments of the time-series graphs indicate (1) there are obvious periodicities at lag 12, (2) apparent upward or downward trends can be located in few series (Figs. 2 and 3), (3) for series that are not contaminated with outliers and/or jumps, stationarity can be attained by differencing the data at lag 12, and (4) transformation of differenced data is unnecessary. Further tests indicate that the assumption of process linearity can be accepted in general.

As mentioned above, the method adopted to induce stationarity is the first step in the identification process of the Box-Jenkins procedure. By applying the appropriate differencing, a time zeries $\{z_t\}$ can be converted to a stationary series $\{x_t\}$. Our analysis showed that homogeneous nonstationarity of the data can be removed by removing the seasonal component (Figs. 4 and 5). Thus, each of the CCS/M series was differenced at lag 12 prior to the application of the forecasting techniques.

3.2 Application of Box-Jenkins technique

Forecasting models for the selected time series, with varying characteristics of the CCS/M variable, were individually identified using the Box-Jenkins approach. In the following, we briefly discuss the application of the method to Series A (Fig. 1). From the autocorrelation function (Fig. 5) of the differenced series $\{x_t\}$

$$x_t = (1 - B^{12})z_t, \tag{16}$$

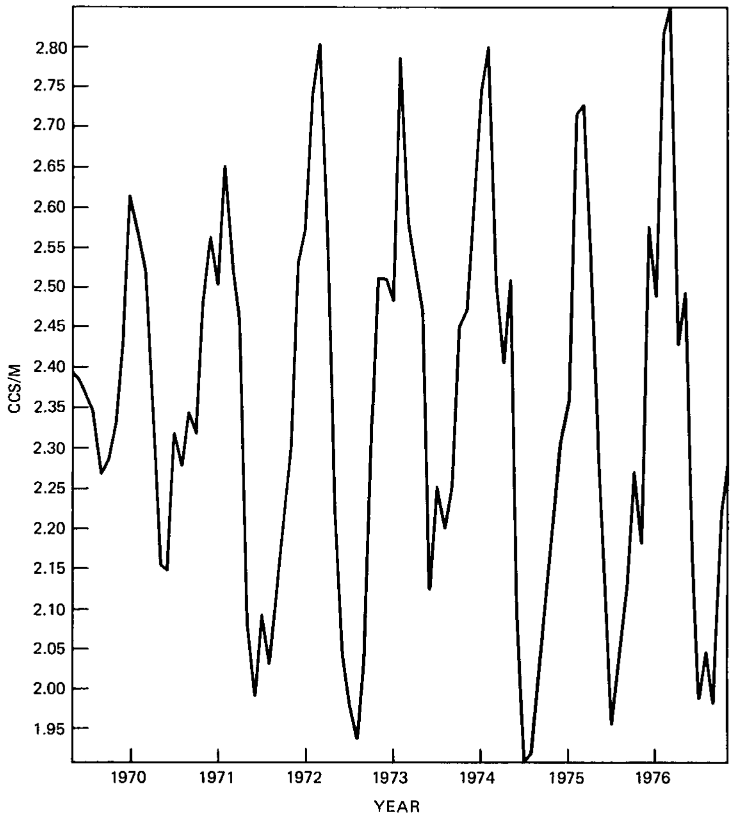

Fig. 1—Monthly data on CCS/M, Series A. Visible seasonal variation—typical behavior.

we note that $\{x_t\}$ is not a white noise process since there is still useful predictive information that can be further explained. As a formal test of hypothesis on the randomness of x_t, we calculated the statistic Q as

$$Q = n \sum_{k=1}^{36} r_k^2(x_t) = 78,$$

where r_k is the kth autocorrelation coefficient of x_t and n (= 60) is the number of points in the differenced series.

For x_t to be white noise, at the 5-percent level

$$Q_{\text{conf}} < \chi^2_{0.95}(36) = 50.$$

Since $Q > Q_{\text{conf}}$, we considered modifications of the model in (16). We note, from the sample autocorrelation function of $\{x_t\}$, that the correlation coefficient at lag 12 is large. This suggests a moving average term of order 12. The autocorrelation pattern suggests a low-order

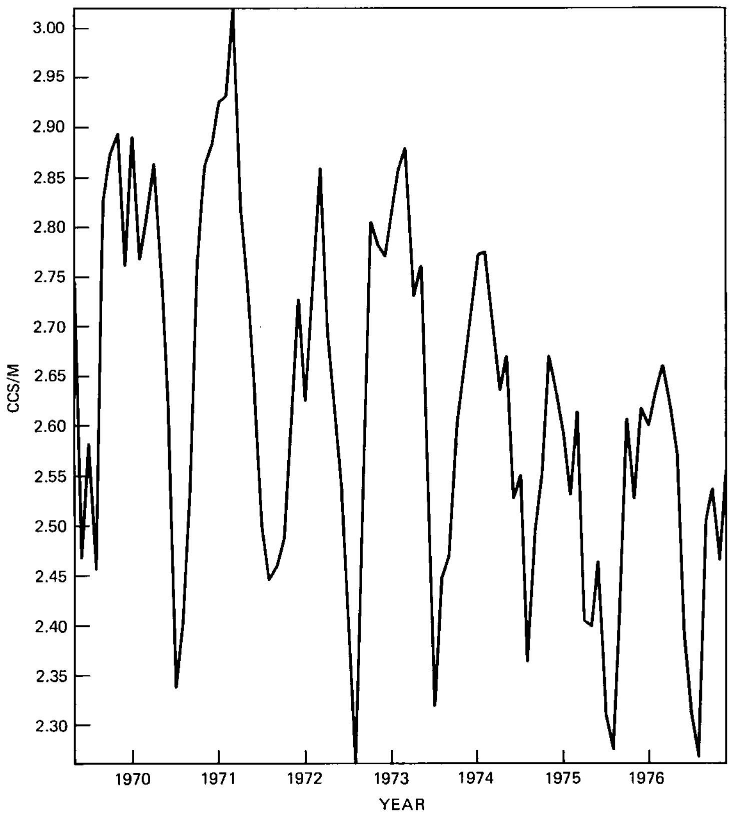

Fig. 2—Monthly data on CCS/M, Series B. Apparent trend variation for peak months—typical behavior.

autoregressive term. Thus we considered the model

$$(1 - \phi B)x_t = (1 - \Theta B^{12})a_t. \qquad (17)$$

After estimation, (17) becomes

$$(1 - 0.63B)(1 - B^{12})z_t = (1 - 0.47B^{12})a_t, \qquad (18)$$

where $\hat{\sigma}_a = 0.12$. Further inspection of the autocorrelation function of the residuals (Fig. 6) suggested no further modification [$Q = x^2(34) = 18$].

3.2.1 *Alternative ARMA models*

Models other than those individually identified by the Box-Jenkins procedure for the examined series were investigated. Alternative models were considered in an attempt to understand the effect of overfitting, underfitting, differencing, and modeling a spurious linear trend on the forecast of CCS/M. For the sake of space, these effects

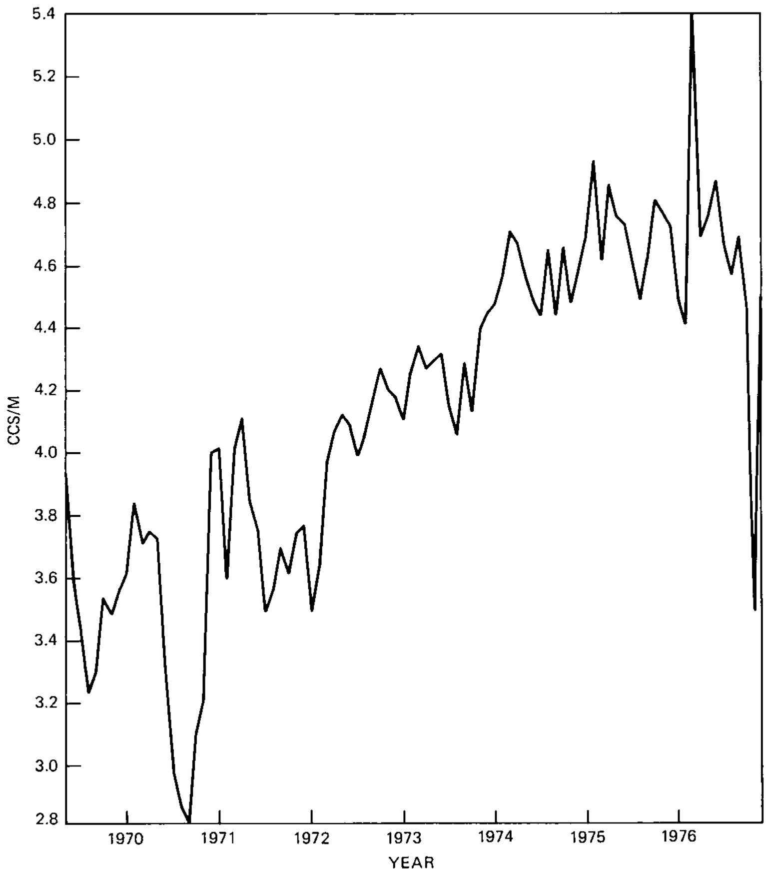

Fig. 3—Monthly data on CCS/M, Series C. Atypical behavior.

will be illustrated graphically by showing the data, and a model fit and its extrapolation, over two to four years for some selected alternatives.

To test the effect of overfitting an MA(1) term to the previously identified form for Series A, consider

$$(1 - \phi B)(1 - B^{12})x_t = (1 - \theta B)(1 - \Theta B^{12})a_t. \qquad (19)$$

The above model specification was examined, since it was identified for several other CCS/M series, using the Box-Jenkins procedure.

After estimation of model parameters, (19) becomes

$$(1 - 0.61B)(1 - B^{12})z_t = (1 + 0.05B)(1 - 0.47B^{12})a_t, \qquad (20)$$

where

$$\hat{\sigma}_a = 0.12, \qquad Q = 19.0.$$

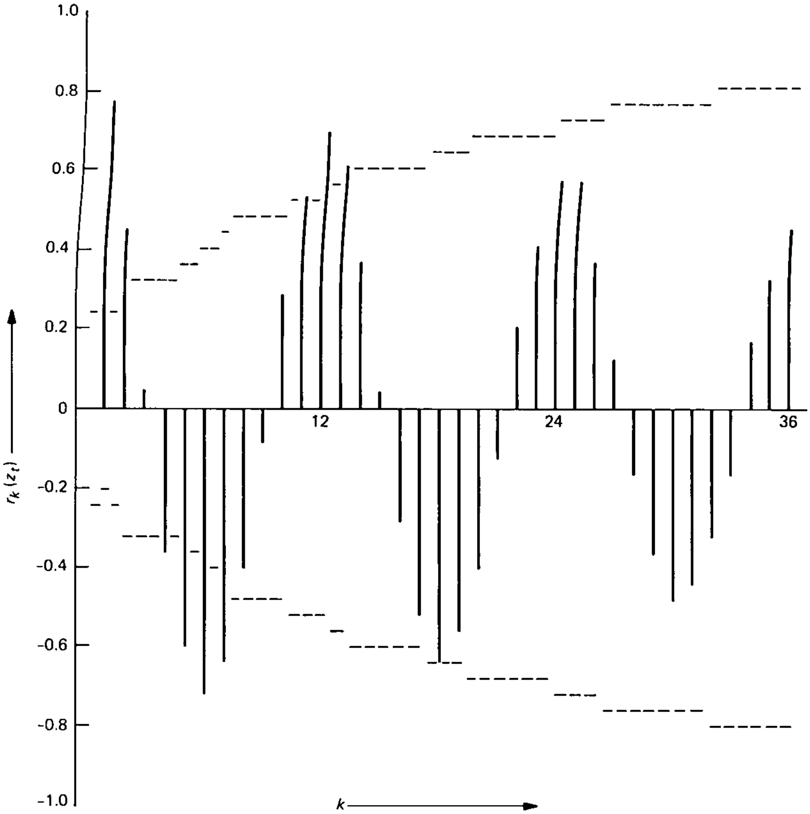

Fig. 4—Estimated autocorrelation, Series A.

The estimated value for the overfitting parameter, θ, is not significantly different from zero, and the goodness-of-fit measure does not indicate inadequacy. And, as expected, this model generates forecasts that are identical to those generated by the original model (18).

Now, consider Series B (Fig. 2). The originally identified Box-Jenkins model for this series is

$$(1 - 0.79B)(1 - B^{12})z_t = (1 - 0.33B)(1 - 0.50B^{12})a_t, \qquad \hat{\sigma}_a = 0.12. \quad (21)$$

Three alternatives are discussed. First

$$(1 - \phi B)(1 - \Phi B^{12})z_t = a_t \quad (22a)$$

with estimated parameters

$$\phi = 0.99, \qquad \Phi = 0.33, \qquad \sigma_a = 0.12.$$

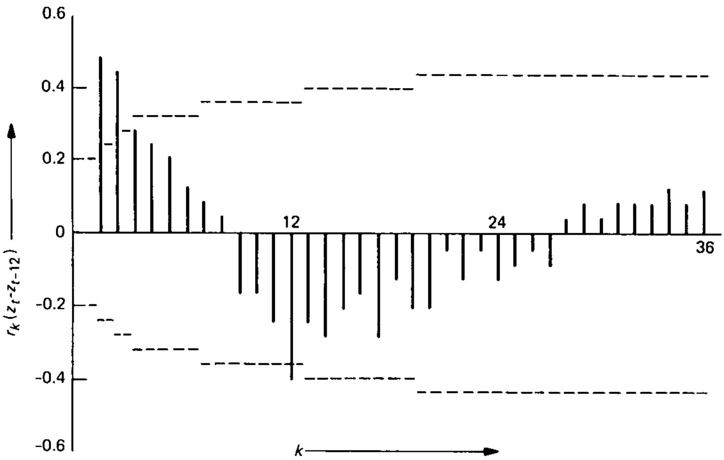

Fig. 5—Estimated autocorrelation, Series A, differenced at Lag 12.

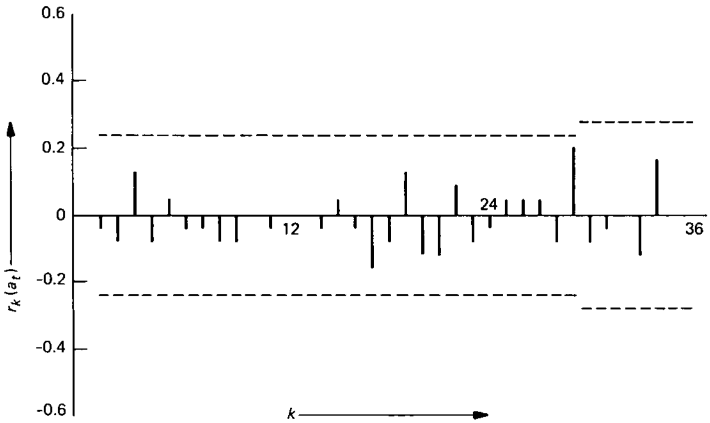

Fig. 6—Estimated autocorrelation of residuals, Series A.

This alternative is considered to determine if the original model, (21), is overdifferencing the time series and, if so, the impact on its forecast.

While both models, (21) and (22a), do not indicate inadequacy from the fitting point of view, their forecasts are markedly different. Figures 7 and 8 give an impression of the forecast performance for these models. It is clear that for model (22a) the seasonality in the forecast damps out quickly. This implies considerable underforecasting of peak months—and consequently of ABS values—for the CCS/M data.

The two other alternatives for Series B were suggested from its graph (Fig. 2). The identified Box-Jenkins model for this series, after differencing for seasonality, represents a stationary process, despite

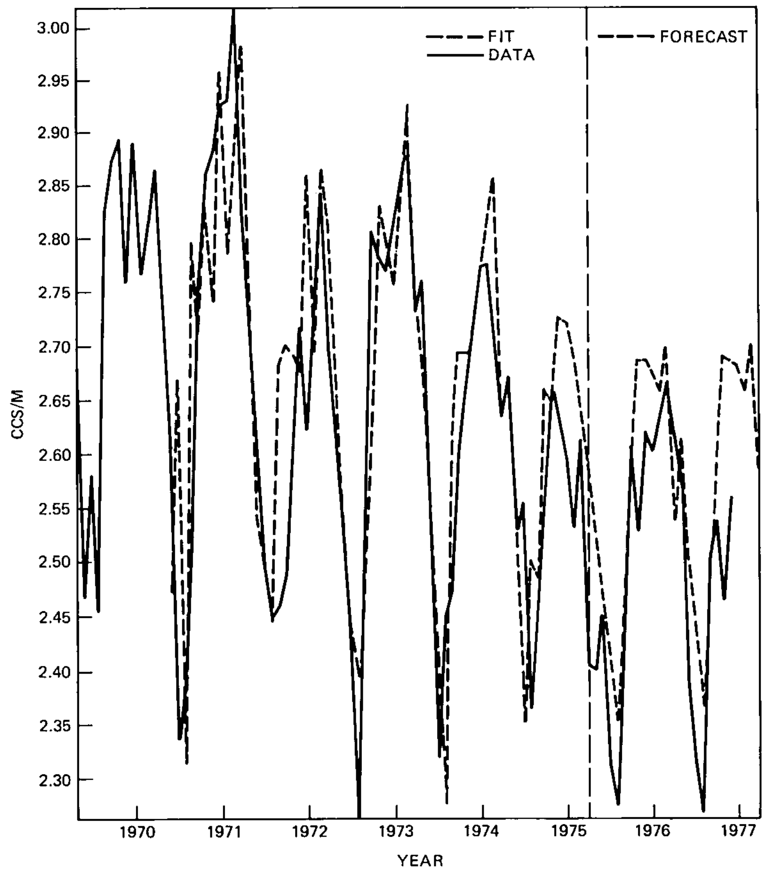

Fig. 7—Data, fitting, and forecast, Series B, model (21).

an apparent linear downward trend. This apparent trend is modeled as short-term fluctuation and is represented by ARMA(1, 1) parameters. If this trend is to be extended into the future, one must consider other representations. For example,

$$(1 - B)(1 - B^{12})z_t = (1 - \theta B)(1 - \Theta B^{12})a_t, \tag{22b}$$

where

$$\hat{\theta} = 0.5, \qquad \hat{\Theta} = 0.53, \qquad \sigma_a = 0.12,$$

and

$$(1 - B^{12})z_t = \theta_0 + (1 - \theta B)(1 - \Theta B^{12})a_t, \tag{22c}$$

where

$$\hat{\theta}_0 = -0.04, \qquad \hat{\theta} = -0.32, \qquad \hat{\Theta} = 0.56, \qquad \sigma_a = 0.12.$$

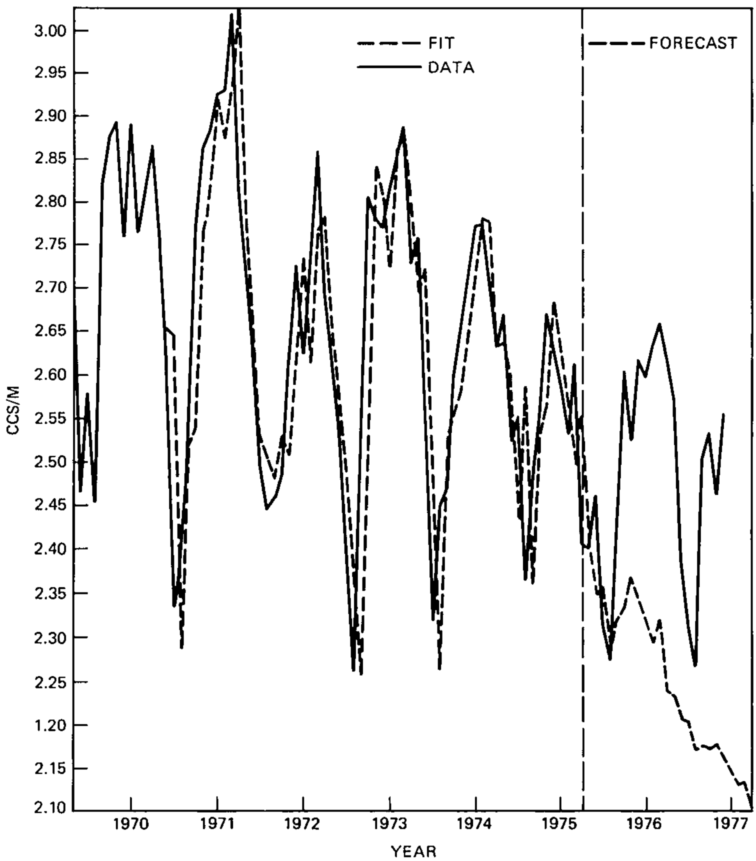

Fig. 8—Data, fitting, and forecast, Series B, model (22a).

Model (22b) simply suggests an exponential smoothing for the year-to-year variation followed by another exponential smoothing for the month-to-month variation. Model (22c) includes a deterministic constant to account for a slope, considered a useful alternative to differencing. It is more restrictive, however, in the sense that it assumes a common slope for the 12 parallel lines (one for each month) connecting different years.

Based on the autocorrelation function of the residuals and on the goodness of fit tests, models (22b) and (22c) may represent Series B. From the forecast point of view, however, the trend effect was much exaggerated by model (22b) and somewhat exaggerated by model (22c) (Figs. 9 and 10). The original model did react to the apparent trend [through the AR(1) term], but was more conservative in its extrapolation (Fig. 7). Because of our experience with the CCS/M data, we favor the original model. We have learned from the behavior of many

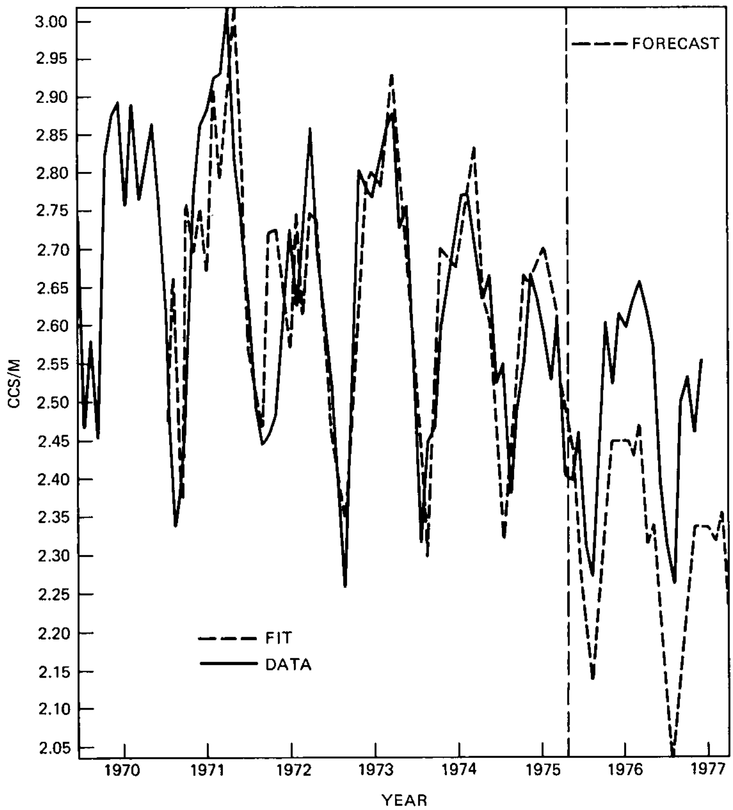

Fig. 9—Data, fitting, and forecast, Series B, model (22b).

time series for CCS/M that a turning point will eventually occur. We emphasize here that one must be careful in allowing for a linear trend representation in a model unless one is certain that this behavior has a physical interpretation. An apparent linear trend may arise simply from the repeated summation of independent disturbances.

3.3 Application of state space approach

Akaike et al. implemented the state space method into a computer package.[8] The original package appeared in 1974 (known as TIMSAC-74) and has been revised several times since. The version we employed here was revised in March 1977.* More recently, the method was implemented by SAS in a procedure called STATESPACE.[15] It is

* A copy of TIMSAC programs was obtained from the Mathematical Sciences Division of the University of Tulsa, Oklahoma. The university distributes these programs for the authors.

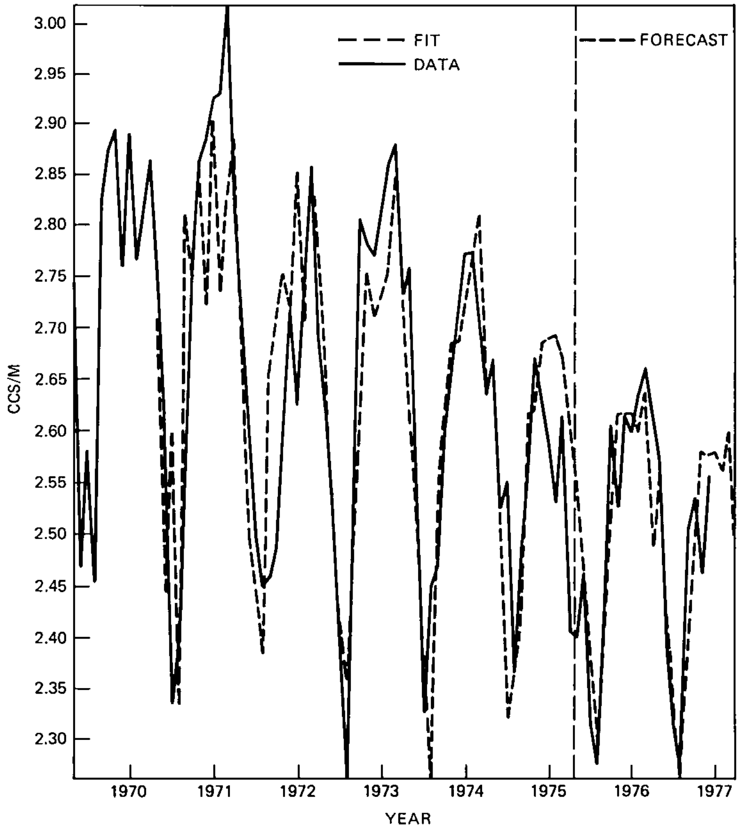

Fig. 10—Data, fitting, and forecast, Series B, model (22c).

assumed that models are limited to the block identifiability condition.[7] In this regard, an identified form using this procedure is equivalent to an ARMA form of order $(p, p - 1)$.

Using TIMSAC programs, models were identified for 15 of the 35 series that were previously identified by the Box-Jenkins approach. While preliminary analysis of these data (Section 3.1) indicated that for the differenced series $\{x_t\}$ there is a dependency between x_t and x_{t-12} (−0.47 for Series A), which suggests a moving average term of order 12, none of the automatically identified models included this term.

We also noted that the numerical procedure for the maximum likelihood computation did not converge easily. It required a large number of iterations. The problem was due to the improper initial values—estimated by the canonical correlation analysis—of model parameters. In many cases, these values specified noninvertible models. For six of the series, numerical breakdowns were encountered as a result of the singularity of the Hessian of the mean log-likelihood function. At the breakdown points, the MAICE model for all series was noise, ARMA(0, 0). As for the other nine series, six were identified as noise [ARMA(0, 0)], two as MA(1) and one as an AR(1). One of the series that was identified as noise is the differenced data for Series A. Neither the sample autocorrelation function of this data (Fig. 4) nor the goodness-of-fit measure is compatible with such representation. We note here that the identified model from the canonical correlation analysis for this series is an AR(1) with $\phi = 0.55$ and AIC = −314. The competing model ARMA(0, 0) was preferred because it has AIC = −316.

3.4 Application of autoregressive approach

Until now, we have been discussing model identification and forecast performance while disregarding the data problems of gross outliers and large jumps. Since these problems can be held accountable for forecast inaccuracies, the idea of eliminating them from the data prior to modeling seemed plausible.

Two procedures for removing jumps were implemented. The first is based on a moving median technique to locate discontinuities in the data; and the other uses a robust regression method to remove the discontinuities found.[16] The latter procedure replaces the time series points before a jump by new values at the most recent level of the process. It also substitutes suspected single outlying points by more probable values.

After modifying the 35 investigated time series using these algorithms, we applied the autoregressive spectrum estimation approach to the cleaned data. A detailed description of this approach and its implementation is given in Ref. 15. Since the considered series are fairly short (less than 250 points),[12] their forecasts were generated from parametric autoregressive models, with their parameters estimated in the frequency domain. The procedure applied here uses the Fast Fourier Transform (FFT) algorithm to estimate the spectrum. The bias that may occur in these estimates due to the limited length of the series was handled by using Tukey's "twicing."[17]

The ABS forecast accuracies obtained using this approach were not significantly different from those found by the random-walk method even for the 1 year ahead, in spite of the screening of the data, the removal of jumps, and the robust estimation of models parameters.

IV. SIMULATIONS

Our objectives in carrying out the simulation study are first to determine whether any of the examined methodology can automatically identify the actual process in an ideal environment. A second objective is to understand the relationship between over- or underparameterization of a stationary process and model adequacy. A third objective is to assess the forecast performance of the various methods and determine if a use of the "best" method is worthwhile. The three simulated processes are

$$\text{Process I} \quad (1 - 0.6B)(1 - B^{12})z_t = (1 - 0.5B^{12})a_t, \quad \sigma_a = 0.12$$

$$\text{Process II} \quad (1 - 0.8B)(1 - B^{12})z_t = (1 - 0.3B)(1 - 0.5B^{12})a_t, \quad \sigma_a = 0.12$$

$$\text{Process III} \quad (1 - B^{12})z_t = a_t, \quad \sigma_a = 0.25.$$

Processes I and II correspond to the two previously identified models for Series A and Series B, respectively. Process III is a seasonal random-walk process (where B is replaced by B^{12}), which was also observed in many of the time-series data.

For each of the three processes, 100 realizations were generated with 252 observations in each. The white noise $\{a_t\}$ was generated from a normal random variate with mean zero and variance equal to those estimated from the actual data. The starting values of these realizations $(z_1, z_2, \cdots, z_{13})$ were taken from the actual CCS/M time series. In all realizations, the first 96 points were discarded, the next 120 points were used for fitting, and the last 36 observations were saved for ex-post analysis. In the following we discuss the simulation results in terms of our three objectives.

4.1 Identification

In this discussion, we exclude the Box-Jenkins approach since its identification stage is based on judgment, while the processes to be identified are known a priori. The state space procedure and the autoregression spectrum estimation procedure were applied to the simulated realizations after differencing at lag 12. Table I gives summary statistics on model identification of time series generated from Processes I and II using the state space procedure. The tabulated statistics are the number of realizations an ARMA(p, q) model is identified, where p and q are given at certain values and/or specified ranges. From these results we note that the procedure did not recognize the distant terms (coefficients of a_{t-12} and a_{t-13}) except for one series. For this series, however, the identified model was ARMA (13, 12), with nonzero estimates for all 25 coefficients. Further inspection of Table I also indicates that the state space procedures did not identify correct parsimonious ARMA models.

We must also indicate that, as with actual data, we encountered the same computational difficulties. These difficulties were, in major part,

Table I—Frequency distribution of the order of the fitted process using Akaike state space procedures

Fitted Order	Frequency	
	Process I	Process II
$p = q = 0$	19	14
$p = 1; q = 0$	16	9
$p = 1; q = 1$	10	18
$p = 0; q = 1$	0	0
$p = 2; q = 1$	15	8
$p = 2; q = 2$	2	8
$p = 0,1; q = 2$	2	2
$3 \leq \max(p, q) \leq 5$	26	29
$6 \leq \max(p, q) \leq 11$	6	8
$\max(p, q) = 12$	3	4
$\max(p, q) = 13$	1	0
Total	100	100

connected with the numerical approximations for the maximization of the likelihood function. The singularity of the Hessian was due to improper assumed values for model parameters. A large number of iterations was required for convergence in most cases.

On the autoregressive spectrum estimation procedure, Table II presents the goodness of fit of its models to the simulated realizations. Since the procedure assumes pure autoregressive models, we will not discuss whether the assumed order over- or underparameterizes a given series. Instead, we compare the estimated value for the standard error of the residual ($\hat{\sigma}_{AS}$ in Table II) for each of the simulated series with the true value of its process. An average estimate of 0.13 (compared to $\sigma_e = 0.12$) for the standard error for Processes I and II realizations indicates that the assumed models may account for most of the predictive information in the simulated data.

4.2 Effect of overparameterizations

To demonstrate the effect of overparameterization, we used (19), denoted hereafter by S, to model each of the 300 simulated series. This model specification was developed as a result of Box-Jenkins analysis of actual CCS/M time series. Table II tabulates the estimated values for the associated parameters ($\hat{\sigma}_S$, $\hat{\phi}_1$, $\hat{\theta}_1$, $\hat{\Theta}_{12}$). Inspection of the estimated standard error ($\hat{\sigma}_S$) indicates that these models can account for the predictive variabilities of the time series (average $\hat{\sigma}_e = 0.12$ for Processes I and II). To compare the estimated values of model coefficients to their true values, however, one must understand the redundancy between autoregressive and moving average terms in a mixed model. For simplicity, consider the form

$$(1 - \phi B)y_t = (1 - \theta B)a_t. \tag{23}$$

Table IIa—Standard errors for AS, RW, and S models and estimates of the S model parameters, with $\sigma_e = 0.12$—Process I realizations

Series	$\hat{\sigma}_{AS}$	$\hat{\sigma}_{RW}$	$\hat{\sigma}_S$	$\hat{\phi}$	$\hat{\theta}$	$\hat{\Theta}$	Series	$\hat{\sigma}_{AS}$	$\hat{\sigma}_{RW}$	$\hat{\sigma}_S$	$\hat{\phi}$	$\hat{\theta}$	$\hat{\Theta}$
A001	0.132	0.158	0.121	0.59	0.00	0.52	A051	0.124	0.168	0.126	0.65	0.07	0.41
02	0.120	0.163	0.114	0.51	−0.29	0.40	052	0.128	0.156	0.122	0.60	0.05	0.41
03	0.126	0.154	0.124	0.39	−0.20	0.32	053	0.117	0.149	0.104	0.53	−0.06	0.59
04	0.123	0.153	0.107	0.53	−0.13	0.55	054	0.117	0.159	0.113	0.68	0.15	0.47
05	0.123	0.189	0.122	0.74	−0.03	0.34	055	0.123	0.167	0.121	0.57	−0.09	0.48
06	0.138	0.171	0.135	0.54	−0.06	0.49	056	0.143	0.196	0.131	0.65	−0.15	0.51
07	0.124	0.143	0.120	0.58	0.05	0.43	057	0.145	0.171	0.124	0.59	0.04	0.79
08	0.137	0.180	0.126	0.84	0.31	0.64	058	0.130	0.161	0.131	0.52	0.01	0.41
09	0.137	0.179	0.131	0.63	0.00	0.52	059	0.143	0.188	0.136	0.65	−0.04	0.54
010	0.133	0.169	0.129	0.71	0.14	0.39	060	0.126	0.159	0.116	0.69	0.15	0.54
011	0.144	0.173	0.128	0.63	0.07	0.59	061	0.145	0.159	0.129	0.63	0.12	0.64
012	0.120	0.165	0.116	0.68	0.06	0.42	062	0.135	0.183	0.127	0.69	0.14	0.48
013	0.142	0.193	0.130	0.68	0.08	0.60	063	0.111	0.155	0.115	0.57	−0.14	0.18
014	0.128	0.175	0.132	0.54	−0.04	0.47	064	0.139	0.173	0.130	0.73	0.25	0.53
015	0.118	0.142	0.113	0.33	−0.25	0.43	065	0.148	0.177	0.128	0.48	−0.13	0.55
016	0.131	0.159	0.117	0.55	−0.08	0.57	066	0.138	0.175	0.125	0.62	0.12	0.57
017	0.120	0.171	0.119	0.70	−0.03	0.42	067	0.142	0.179	0.127	0.60	−0.06	0.59
018	0.136	0.165	0.134	0.47	−0.06	0.54	068	0.138	0.178	0.117	0.48	−0.31	0.64
019	0.123	0.173	0.119	0.74	0.16	0.47	069	0.140	0.204	0.137	0.63	−0.04	0.50
020	0.126	0.166	0.124	0.63	0.09	0.48	070	0.123	0.145	0.121	0.55	0.04	0.38
021	0.131	0.171	0.126	0.57	−0.11	0.39	071	0.136	0.177	0.130	0.72	0.23	0.43
022	0.132	0.167	0.125	0.63	0.04	0.45	072	0.130	0.156	0.129	0.63	0.20	0.45
023	0.125	0.171	0.122	0.72	0.35	0.44	073	0.126	0.172	0.124	0.77	0.14	0.38
024	0.138	0.174	0.124	0.79	0.29	0.56	074	0.126	0.143	0.120	0.68	0.20	0.44
025	0.130	0.163	0.125	0.43	−0.23	0.39	075	0.116	0.165	0.114	0.48	−0.36	0.45
026	0.103	0.154	0.113	0.69	0.13	0.29	076	0.146	0.193	0.128	0.83	0.22	0.67
027	0.127	0.182	0.120	0.68	−0.02	0.53	077	0.137	0.179	0.133	0.61	−0.04	0.55
028	0.121	0.151	0.120	0.76	0.31	0.37	078	0.122	0.154	0.121	0.38	−0.20	0.51
029	0.120	0.134	0.113	0.38	−0.15	0.36	079	0.126	0.150	0.114	0.57	0.08	0.51
030	0.127	0.158	0.121	0.43	−0.06	0.52	080	0.132	0.169	0.121	0.63	−0.08	0.49
031	0.115	0.151	0.115	0.66	−0.09	0.32	081	0.132	0.171	0.129	0.70	0.11	0.47
032	0.124	0.163	0.118	0.70	0.01	0.52	082	0.127	0.170	0.116	0.66	0.03	0.43
033	0.133	0.182	0.133	0.81	0.32	0.31	083	0.140	0.170	0.128	0.39	−0.17	0.51
034	0.139	0.179	0.134	0.58	0.01	0.57	084	0.116	0.139	0.116	0.48	−0.11	0.40
035	0.147	0.191	0.130	0.65	−0.01	0.54	085	0.127	0.174	0.117	0.61	−0.02	0.63
036	0.122	0.148	0.110	0.58	−0.01	0.45	086	0.115	0.167	0.119	0.50	−0.30	0.35
037	0.125	0.153	0.112	0.44	−0.15	0.47	087	0.125	0.154	0.115	0.25	−0.48	0.54
038	0.123	0.150	0.118	0.32	−0.20	0.46	088	0.140	0.167	0.123	0.76	0.31	0.51
039	0.114	0.140	0.120	0.50	0.16	0.41	089	0.125	0.168	0.110	0.75	0.04	0.51
040	0.136	0.188	0.123	0.70	0.02	0.68	090	0.111	0.133	0.106	0.69	0.26	0.38
041	0.132	0.196	0.116	0.66	−0.09	0.51	091	0.121	0.174	0.123	0.73	0.13	0.24
042	0.113	0.140	0.115	0.46	−0.14	0.29	092	0.128	0.160	0.114	0.74	0.19	0.63
043	0.131	0.185	0.129	0.77	0.20	0.38	093	0.133	0.175	0.129	0.62	−0.08	0.37
044	0.146	0.185	0.124	0.57	−0.11	0.67	094	0.122	0.164	0.115	0.61	−0.02	0.55
045	0.136	0.171	0.129	0.50	−0.02	0.60	095	0.130	0.174	0.117	0.72	0.11	0.51
046	0.132	0.166	0.117	0.62	−0.01	0.61	096	0.137	0.187	0.125	0.64	−0.01	0.46
047	0.130	0.167	0.126	0.70	0.17	0.48	097	0.121	0.162	0.113	0.61	−0.12	0.48
048	0.138	0.172	0.130	0.43	−0.15	0.46	098	0.140	0.187	0.128	0.60	−0.07	0.55
049	0.135	0.188	0.129	0.74	0.11	0.46	099	0.137	0.174	0.134	0.50	−0.01	0.46
050	0.132	0.173	0.119	0.60	0.05	0.57	100	0.119	0.157	0.116	0.51	−0.10	0.44

This can be written as

$$y_t = [1 + (\phi - \theta)B + (\phi^2 - \theta\phi)B^2 + (\phi^3 - \theta\phi^2)B^3 + \cdots]a_t. \quad (24)$$

For $\phi = \theta = 0$, (23) represents a noise process. But when $\phi = \theta \neq 0$ it still represents a noise process, since all terms of the polynomial (24) cancel out.

For the model specification (19), similar redundancy exists between the AR(1) coefficient, ϕ, and the MA(1) coefficient, θ, and to a lesser degree it exists between ϕ and the moving average coefficients at lags 12 and 13. The effect of these redundancies is particularly clear from the estimated coefficients of Process III realizations ($\hat{\phi} = \hat{\theta} \neq 0$).

Table IIb—Standard errors for AS, RW, and S models and estimates of the S model parameters, with $\sigma_e = 0.12$—Process II realizations

Series	$\hat{\sigma}_{AS}$	$\hat{\sigma}_{RW}$	$\hat{\sigma}_S$	$\hat{\phi}$	$\hat{\theta}$	$\hat{\Theta}$	Series	$\hat{\sigma}_{AS}$	$\hat{\sigma}_{RW}$	$\hat{\sigma}_S$	$\hat{\phi}$	$\hat{\theta}$	Θ
B001	0.132	0.156	0.122	0.77	0.28	0.52	B051	0.124	0.175	0.126	0.80	0.28	0.42
02	0.120	0.165	0.114	0.73	0.06	0.43	052	0.128	0.162	0.122	0.83	0.38	0.41
03	0.126	0.159	0.125	0.76	0.32	0.33	053	0.116	0.149	0.104	0.65	0.15	0.59
04	0.123	0.155	0.107	0.81	0.29	0.55	054	0.118	0.165	0.113	0.85	0.42	0.48
05	0.124	0.197	0.122	0.85	0.18	0.35	055	0.124	0.171	0.120	0.87	0.40	0.52
06	0.138	0.175	0.134	0.70	0.19	0.51	056	0.143	0.207	0.131	0.84	0.16	0.53
07	0.124	0.145	0.120	0.85	0.43	0.43	057	0.145	0.175	0.124	0.86	0.44	0.78
08	0.137	0.190	0.127	0.92	0.46	0.65	058	0.130	0.161	0.131	0.76	0.35	0.42
09	0.137	0.178	0.131	0.81	0.28	0.52	059	0.143	0.197	0.136	0.87	0.30	0.56
10	0.134	0.177	0.129	0.90	0.46	0.42	060	0.126	0.169	0.115	0.85	0.40	0.54
011	0.144	0.177	0.127	0.91	0.52	0.62	061	0.146	0.162	0.128	0.90	0.54	0.47
012	0.120	0.172	0.116	0.79	0.24	0.44	062	0.135	0.191	0.127	0.81	0.33	0.49
013	0.142	0.201	0.130	0.83	0.31	0.61	063	0.110	0.154	0.115	0.71	0.11	0.19
014	0.129	0.178	0.132	0.71	0.22	0.47	064	0.139	0.176	0.131	0.86	0.43	0.52
015	0.117	0.135	0.111	0.44	−0.01	0.44	065	0.148	0.172	0.128	0.73	0.22	0.56
016	0.130	0.157	0.117	0.81	0.30	0.57	066	0.138	0.179	0.126	0.75	0.33	0.60
017	0.120	0.157	0.119	0.85	0.22	0.43	067	0.142	0.117	0.127	0.73	0.16	0.57
018	0.136	0.160	0.133	0.60	0.16	0.55	068	0.138	0.176	0.117	0.65	−0.07	0.64
019	0.123	0.184	0.120	0.84	0.33	0.48	069	0.139	0.207	0.137	0.76	0.17	0.50
020	0.126	0.170	0.124	0.77	0.32	0.48	070	0.123	0.145	0.121	0.73	0.30	0.38
021	0.131	0.168	0.125	0.71	0.15	0.39	071	0.135	0.182	0.131	0.85	0.41	0.43
022	0.132	0.169	0.125	0.81	0.31	0.45	072	0.130	0.159	0.129	0.84	0.48	0.47
023	0.125	0.149	0.120	0.93	0.66	0.46	073	0.124	0.184	0.124	0.88	0.34	0.38
024	0.138	0.189	0.125	0.90	0.45	0.53	074	0.127	0.150	0.120	0.87	0.47	0.44
025	0.130	0.159	0.125	0.62	0.07	0.39	075	0.116	0.167	0.115	0.64	−0.09	0.44
026	0.104	0.162	0.115	0.83	0.31	0.29	076	0.147	0.208	0.127	0.94	0.42	0.71
027	0.127	0.190	0.119	0.88	0.34	0.55	077	0.137	0.189	0.133	0.84	0.31	0.55
028	0.121	0.160	0.120	0.90	0.51	0.39	078	0.122	0.152	0.120	0.80	0.41	0.51
029	0.120	0.135	0.113	0.77	0.37	0.35	079	0.127	0.155	0.113	0.80	0.40	0.52
030	0.127	0.161	0.123	0.62	0.20	0.51	080	0.132	0.170	0.121	0.78	0.16	0.49
031	0.115	0.163	0.115	0.85	0.40	0.33	081	0.132	0.178	0.129	0.80	0.30	0.49
032	0.124	0.166	0.118	0.86	0.28	0.52	082	0.127	0.177	0.117	0.81	0.26	0.43
033	0.134	0.204	0.132	0.92	0.49	0.35	083	0.139	0.168	0.128	0.71	0.29	0.50
034	0.138	0.179	0.133	0.88	0.46	0.61	084	0.118	0.138	0.116	0.70	0.22	0.40
035	0.147	0.201	0.130	0.79	0.23	0.55	085	0.126	0.173	0.117	0.75	0.21	0.63
036	0.122	0.143	0.109	0.67	0.18	0.45	086	0.114	0.171	0.119	0.72	0.08	0.39
037	0.124	0.148	0.112	0.61	0.14	0.48	087	0.124	0.145	0.116	0.42	−0.17	0.51
038	0.122	0.141	0.118	0.53	0.11	0.43	088	0.140	0.171	0.125	0.85	0.45	0.50
039	0.114	0.139	0.120	0.77	0.53	0.40	089	0.125	0.177	0.111	0.87	0.25	0.51
040	0.136	0.198	0.123	0.85	0.26	0.68	090	0.111	0.143	0.106	0.83	0.45	0.37
041	0.132	0.204	0.116	0.81	0.15	0.51	091	0.121	0.177	0.122	0.81	0.31	0.26
042	0.113	0.138	0.115	0.66	0.14	0.28	092	0.128	0.166	0.114	0.80	0.38	0.63
043	0.131	0.199	0.123	0.87	0.38	0.39	093	0.133	0.175	0.129	0.74	0.12	0.36
044	0.146	0.188	0.125	0.81	0.25	0.67	094	0.121	0.171	0.115	0.75	0.21	0.56
045	0.136	0.168	0.128	0.63	0.19	0.60	095	0.130	0.179	0.116	0.88	0.40	0.53
046	0.132	0.168	0.117	0.76	0.23	0.60	096	0.137	0.187	0.124	0.74	0.19	0.46
047	0.131	0.176	0.127	0.83	0.36	0.50	097	0.121	0.159	0.113	0.71	0.08	0.46
048	0.138	0.174	0.131	0.83	0.44	0.48	098	0.140	0.189	0.128	0.76	0.19	0.55
049	0.134	0.200	0.129	0.89	0.35	0.47	099	0.135	0.173	0.133	0.65	0.26	0.46
050	0.132	0.181	0.119	0.78	0.32	0.57	100	0.120	0.162	0.116	0.82	0.36	0.43

Based on these results, we conclude that (19) can represent the generated processes adequately. In other words, overparameterization by a moving average term at lag 1, as in Process I, and furthermore by an autoregressive term at lag 1 and moving average terms at lags 12 and 13, as in Process III, did not affect the standard error of the residuals significantly.

4.3 Forecast accuracy

For this evaluation, we consider forecasts generated by: (a) the model specification in (19) (S), (b) the autoregressive spectral esti-

Table IIc—Standard errors for AS, RW, and S models and estimates of the S model parameters, with $\sigma_e = 0.25$—Process III realizations

Series	$\hat{\sigma}_{AS}$	$\hat{\sigma}_{RW}$	$\hat{\sigma}_{S}$	$\hat{\phi}$	$\hat{\theta}$	$\hat{\Theta}$	Series	$\hat{\sigma}_{AS}$	$\hat{\sigma}_{RW}$	$\hat{\sigma}_{S}$	$\hat{\phi}$	$\hat{\theta}$	$\hat{\Theta}$
C001	0.239	0.251	0.242	−0.63	−0.75	0.08	C051	0.236	0.251	0.254	0.11	0.11	−0.04
02	0.226	0.229	0.226	−0.25	−0.48	−0.07	052	0.244	0.253	0.252	−0.24	−0.22	−0.07
03	0.250	0.263	0.258	0.87	0.94	−0.17	053	0.207	0.209	0.208	−0.80	−0.79	0.17
04	0.227	0.223					054	0.215	0.226	0.227	0.02	0.09	0.06
05	0.242	0.256	0.252	0.66	0.51	−0.17	055	0.229	0.246	0.249	0.08	0.00	−0.02
06	0.263	0.275	0.275	0.89	0.98	−0.08	056	0.260	0.270	0.270	−0.05	−0.25	0.00
07	0.241	0.245	0.248	−0.63	−0.59	−0.06	057	0.261	0.264	0.266	−0.30	−0.30	0.16
08	0.254	0.261	0.264	−0.75	−0.67	−0.01	058	0.242	0.261	0.263	0.00	0.04	−0.00
09	0.255	0.260	0.262	0.90	0.87	0.10	059	0.269	0.273	0.276	0.00	−0.12	−0.01
010	0.260	0.264	0.266	−0.68	−0.65	−0.16	060	0.225	0.231	0.234	0.45	0.44	0.09
011	0.255	0.257	0.259	−0.56	−0.52	0.13	061	0.267	0.266	0.269	0.23	0.30	0.01
012	0.239	0.236	0.239	0.49	0.45	−0.10	062	0.248	0.254	0.258	0.32	0.35	0.03
013	0.253	0.265	0.267	−0.46	−0.40	0.14	063	0.232	0.238	0.233	−0.07	−0.20	−0.28
014	0.239	0.255	0.257	−0.65	−0.57	−0.02	064	0.253	0.266	0.268	0.78	0.79	0.01
015	0.223	0.234	0.230	0.09	0.06	0.05	065	0.258	0.262	0.257	0.57	0.65	0.18
016	0.239	0.236	0.237	−0.56	−0.63	0.13	066	0.246	0.256	0.252	−0.76	−0.61	0.13
017	0.234	0.240	0.240	0.60	0.45	−0.05	067	0.254	0.255	0.258	0.43	0.36	0.13
018	0.257	0.277	0.270	−0.88	−0.92	−0.04	068	0.240	0.248	0.244	−0.07	−0.29	0.19
019	0.226	0.245	0.243	0.77	0.77	−0.00	069	0.267	0.274	0.277	0.66	0.58	−0.09
020	0.242	0.248	0.251	−0.86	−0.81	−0.07	070	0.234	0.249				
021	0.256	0.257					071	0.257	0.265	0.266	−0.67	−0.53	−0.04
022	0.247	0.254	0.257	0.40	0.36	−0.04	072	0.241	0.257	0.254	−0.69	−0.50	0.00
023	0.236	0.252	0.252	−0.64	−0.49	0.01	073	0.244	0.253	0.254	0.82	0.75	−0.13
024	0.261	0.257	0.261	−0.24	−0.17	−0.03	074	0.241	0.244	0.244	−0.45	−0.31	−0.01
025	0.256	0.263	0.265	0.19	0.12	−0.08	075	0.220	0.246	0.239	−0.08	−0.37	−0.06
026	0.209	0.236	0.227	−0.46	−0.38	−0.16	076	0.267	0.267	0.264	0.91	0.86	0.25
027	0.228	0.239	0.240	−0.17	−0.29	0.09	077	0.256	0.267	0.268	0.84	0.82	0.01
028	0.232	0.243	0.242	−0.60	−0.43	−0.05	078	0.225	0.245	0.245	0.47	0.57	0.01
029	0.234	0.237	0.239	−0.45	−0.41	−0.07	079	0.233	0.234	0.233	0.28	0.38	−0.00
030	0.239	0.247	0.243	−0.24	−0.17	0.02	080	0.240	0.242	0.239	−0.22	−0.40	0.08
031	0.239	0.241	0.238	0.04	0.01	−0.18	081	0.251	0.263	0.267	0.82	0.80	−0.09
032	0.235	0.243	0.246	0.31	0.23	−0.01	082	0.240	0.244	0.249	0.09	0.08	−0.01
033	0.253	0.263	0.264	0.92	0.86	−0.04	083	0.266	0.266	0.270	−0.73	−0.78	−0.05
034	0.246	0.267	0.267	0.49	0.52	0.09	084	0.231	0.232	0.235	0.76	0.83	−0.04
035	0.249	0.259	0.259	0.44	0.37	0.18	085	0.222	0.239	0.241	−0.09	−0.08	0.16
036	0.227	0.238	0.234	0.09	0.07	−0.02	086	0.239	0.242	0.239	0.24	0.08	−0.06
037	0.227	0.229	0.232	−0.68	−0.68	0.05	087	0.230	0.242	0.239	−0.62	−0.82	0.02
038	0.231	0.242	0.241	0.72	0.80	0.02	088	0.257	0.260	0.258	−0.29	−0.10	0.06
039	0.216	0.244	0.237	0.32	0.53	−0.09	089	0.234	0.232	0.232	0.70	0.56	0.03
040	0.247	0.253	0.254	0.81	0.73	0.09	090	0.220	0.220	0.219	0.03	0.17	−0.12
041	0.234	0.240	0.240	0.40	0.25	0.07	091	0.243	0.245	0.238	0.76	0.79	−0.14
042	0.228	0.231	0.231	0.20	0.18	−0.20	092	0.228	0.231	0.229	−0.85	−0.72	0.13
043	0.267	0.275	0.276	0.42	0.42	−0.13	093	0.251	0.260	0.262	0.38	0.28	−0.09
044	0.254	0.251	0.245	−0.91	−0.99	0.21	094	0.227	0.238	0.238	0.80	0.82	0.01
045	0.239	0.262	0.262	0.32	0.39	0.16	095	0.231	0.235	0.226	1.00	1.05	0.11
046	0.241	0.234	0.237	−0.75	−0.81	0.10	096	0.260	0.260	0.264	0.14	0.10	−0.03
047	0.246	0.258	0.261	−0.17	−0.07	0.01	097	0.227	0.230	0.232	0.18	0.03	−0.03
048	0.254	0.259	0.255	−0.96	−1.02	−0.03	098	0.258	0.260	0.264	−0.09	−0.14	0.05
049	0.247	0.266	0.265	0.92	0.88	−0.02	099	0.255	0.266	0.270	−0.56	−0.53	−0.01
050	0.246	0.245	0.249	0.08	0.09	0.02	100	0.231	0.233	0.236	−0.09	−0.10	−0.06

mation method (AS) and, (c) the random-walk model (RW) for the simulated series.

While minimization of the mean square errors during the model-fitting phase guarantees minimum mean square errors for the future of the simulated data (post-sample phase), we consider ex-post analysis to assess the effect of conversion of the monthly forecast to ABS values on forecast accuracy.

The ABS quantity is made up of the average of the three highest months in a year. The busy season months can vary from one year to the next. A nonbusy season month of one year may become a busy season month of the next year as a result of some added white noise.

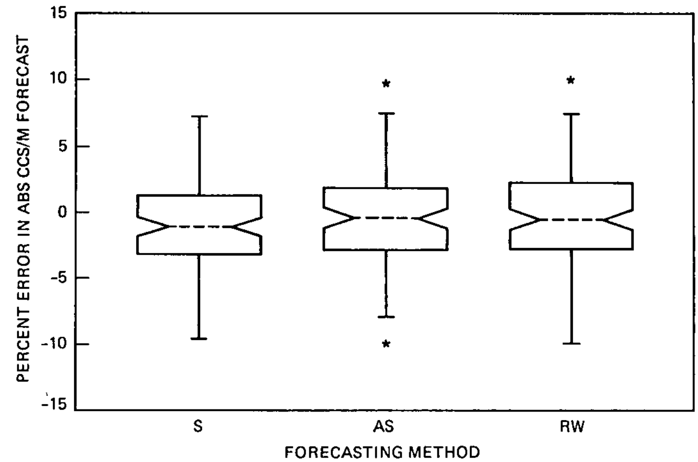

Fig. 11—Notched box plot for one-year-ahead forecast error—Process I realizations.

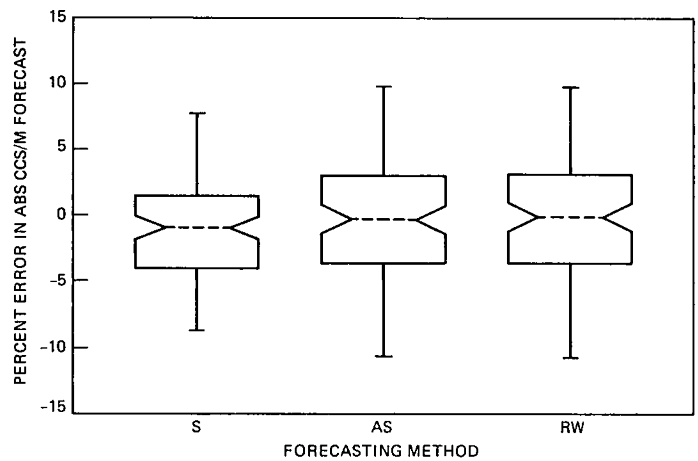

Fig. 12—Notched box plot for two-year-ahead forecast error—Process I realizations.

Since the effect of this aggregation is not a simple analytical problem, we resort to ex-post analysis of the simulated realizations for explanation.

Figures 11 through 16 present these comparisons in the form of box plots* of the percent forecast accuracy[18] for series that are grouped by

*In a box plot the middle line is the median and the upper and lower lines of the box are the upper and lower quartiles, respectively. The box whiskers are drawn out to the nearest values within 1.5 quartiles. Points outside the whiskers are considered outliers and are indicated by asterisks. The notches in the side of a box represents a rough 95-percent confidence interval for the median.

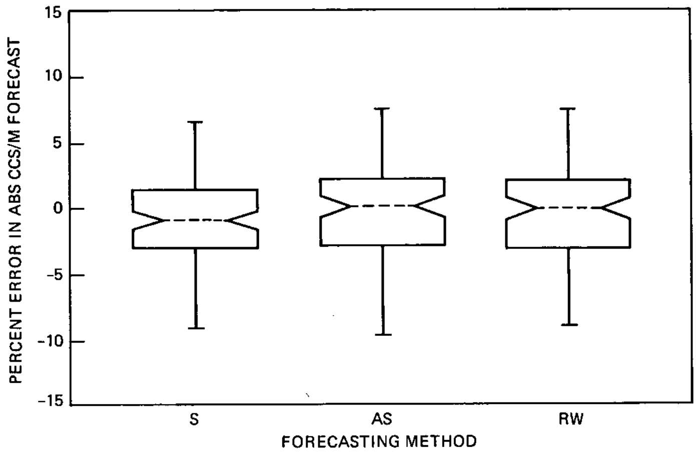

Fig. 13—Notched box plot for one-year-ahead forecast error—Process II realizations.

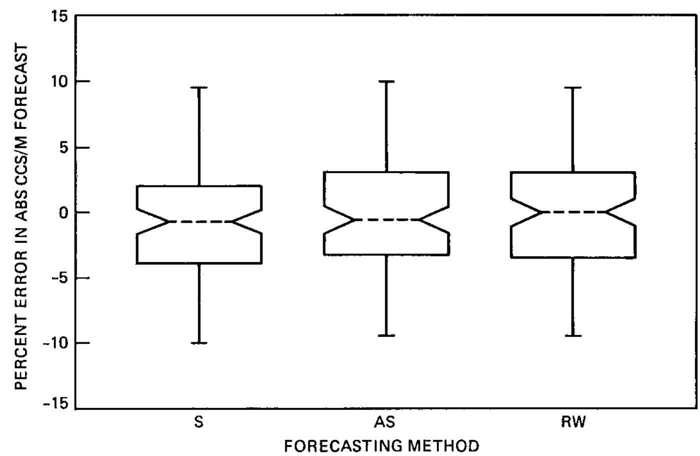

Fig. 14—Notched box plot for two-year-ahead forecast error—Process II realizations.

their process type for one and two years ahead. Based on these graphs, we draw a general conclusion that there are no significant differences in the forecasting accuracies among the three methods.

We did notice, however, that an optimal model that represents the actual process, and is expected to provide best forecasts, results in low prediction of the ABS values. More specifically, the S form is the model specification for Process II. Thus, it is expected to generate "best" forecasts. Figures 13 and 14, however, indicate a bias towards underforecast of ABS values for optimal models. The estimated medians of the percent errors for one- and two-years-ahead forecasts are

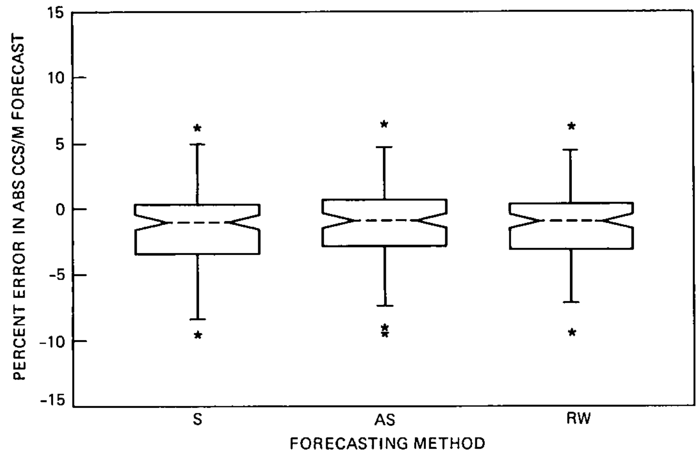

Fig. 15—Notched box plot for one-year-ahead forecast error—Process III realizations.

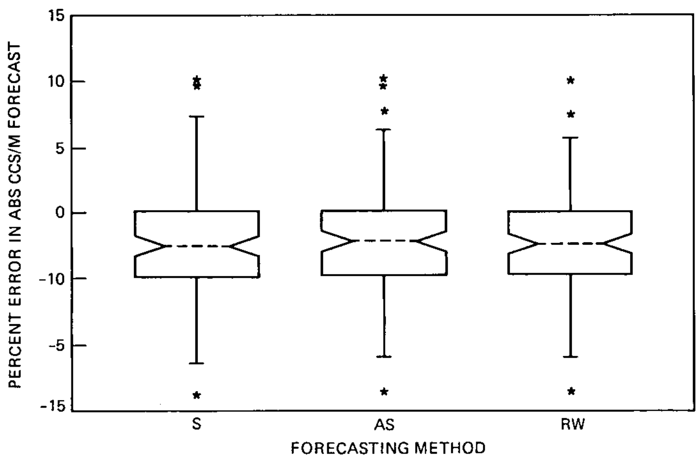

Fig. 16—Notched box plot for two-year-ahead forecast error—Process III realizations.

−1.0 and −0.7, respectively. This observation is also repeated for Process III. The random-walk model underforecasts ABS values for realizations generated by the random-walk process. Figures 15 and 16 demonstrate this observation. The estimated medians for one and two years ahead were −1.0 and −2.4, respectively (notice that here $\sigma_a = 0.25$). In this regard, the ABS forecast is not optimal.

In the following we give a heuristic argument on the above observation and explain why the ABS forecast is biased. When the ABS forecast is made at time t, ($A\hat{B}S_t$), it is conventionally derived from the monthly forecast by

Table III—Simulation results. Frequency method A outperforms Method B. There are 100 realizations generated from each process

Method A	Process No.	Method B: 1 Year Ahead S	AS	RW	2 Years Ahead S	AS	RW	3 Years Ahead S	AS	RW
S		—	57	52	—	62*	58	—	50	47
AS	I	43	—	45	37	—	40	49	—	39
RW		45	40	—	39	45	—	51	43	—
S		—	51	51	—	58	52	—	50	43
AS	II	46	—	45	39	—	42	50	—	43
RW		45	38	—	45	41	—	54	41	—
S		—	41	32	—	52	36	—	39	42
AS	III	50	—	41	42	—	41	45	—	45
RW		46	48	—	44	43	—	44	41	—

* Indicates significance at 5-percent level.

$$A\hat{B}S_t = ABS(\hat{z}_t, \hat{z}_{t+1}, \cdots, \hat{z}_{t+12}).$$

Suppose z_t is white noise and the maximum forecast for the next k months is required. The best forecast is the expected value of the maximum of k-order statistics from the normal distribution, which is a positive number. On the other hand, the maximum of the predictions of $z_{t+1}, z_{t+2}, \cdots, z_{t+k}$ is zero. This observation indicates the bias in $A\hat{B}S$. However, one can adjust for this bias since it can be estimated directly from the simulation results.

For further evaluation of forecast accuracy, Table III gives the frequency that Method A outperforms Method B. This table was generated using a sign test.[19] This test does not take into account the magnitude of the error.

V. SUMMARY AND CONCLUSIONS

Through extensive data analysis of actual and simulated time series we answered, in this paper, numerous crucial questions: specific questions concerning the selection of a best univariate method for forecasting telephone usage demand, and general questions on the weaknesses and strengths of the three state-of-the-art techniques.

On the specific questions, our results indicate that none of the investigated techniques outperformed a naive random walk in forecasting the ABS CCS/M variable. By means of simulation results we show that there is no gain in using optimal models (models that represent the exact process) for generating point forecasts over a random-walk model for lead times of more than one year. Moreover, we pointed out a bias in the ABS measure which can be adjusted for, using the simulation result.

Concerning the general issues, we demonstrated that identification of the nonstationary behavior is critical to the forecast accuracy. It is this behavior that has a long-lasting effect on a process. While accuracy is sensitive to this identification, proper techniques are lacking. Of the three investigated methods, only the Box-Jenkins approach gives some guidelines, but it requires user's judgment.

The simulation results also indicate that the AIC criterion did not identify parsimonious models for the investigated processes. They point out that the nonconvergence of the numerical procedures for the state space approach can add considerable complexities and can increase computer costs—all of which must be considered when assessing the feasibility of the method as an automatic forecasting technique.

REFERENCES

1. N. D. Blair and W. S. Hayward, Jr., "New Methods for Administrative Area and Local Entity Forecasting," Ninth ITC Conf, Spain, October 1979.
2. G. E. P. Box and G. M. Jenkins, *Time Series Analysis Forecasting and Control*, San Francisco: Holden-Day, 1976, p. 575.
3. H. Akaike, "A New Look at the Statistical Model Identification," IEEE Trans. Automat. Contr., *AC-19* (December 1974), pp. 716–23.
4. R. H. Jones, "Identification and Autoregressive Spectrum Estimation," IEEE Trans. Automat. Contr., *AC-19* (December 1974), pp. 894–8.
5. R. W. Katz, "On Some Criteria for Estimating the Order of a Markov Chain," ASA Annual Meeting, Washington, D.C., August 1979.
6. H. Akaike, "Markovian Representation of Stochastic Processes and its Application to the Analysis of Autoregressive Moving Average Processes," Ann. Inst. Statist. Math., *26*, No. 3 (1974), pp. 363–87.
7. H. Akaike, "Canonical Correlation Analysis of Time Series and the Use of an Information Criterion," in R. K. Mehra and D. G. Lainiotis, eds., *System Identification: Advances and Case Studies*, New York: Academic Press, 1976.
8. H. Akaike, E. Arahata, and T. Ozaki, "TIMSAC-74-A Time Series Analysis and Control Program Package −(1)," Computer Science Monographs, Tokyo, March 1975.
9. E. Parzen, "An Approach to Time Series Analysis," Annals of Mathematical Statistics, *32*, No. 4 (December 1961), pp. 951–89.
10. H. Akaike, "Fitting Autoregressive Models for Prediction," Ann. Inst. Statist. Math., *21*, No. 2 (1969), pp. 243–7.
11. P. Whittle, *Prediction and Regulation by Linear Least Square Methods*, Princeton: D. Van Nostrand, 1963.
12. B. J. Bhansali, "Asymptotic Properties of the Wiener-Kolmogorov Predictor," JRSS, Series B, *36*, No. 1 (1974), pp. 61–73.
13. J. W. Tukey, "An Introduction to the Calculations of Numerical Spectrum Analysis," in B. Harris, ed., *Spectral Analysis of Time Series*, New York: Wiley, 1967.
14. W. M. Brelsford, D. A. Relles, *STATLIB: A Statistical Computing Library*, Englewood Cliffs, NJ: Prentice-Hall, 1981, p. 427.
15. D. DeLong, "A State Space Approach to Modeling Vector Time Series," ASA Annual Meeting, Houston, Texas, August 1981.
16. B. Kleiner, R. D. Martin, and D. J. Thomson, "Robust Estimation of Power Spectra," JRSS, Series B, *41*, No. 3 (1979), pp. 313–51.
17. J. W. Tukey, *Exploratory Data Analysis*, Princeton: Addison Wesley, 1977.
18. R. McGill, J. W. Tukey, and W. A. Larson, "Variations of Box Plots," American Statistician, *32* (February 1978), pp. 12–16.
19. A. M. Mood, "On the Asymptotic Efficiency of Certain Nonparametric Two Sample Tests," Annals of Mathematical Statistics, *25* (September 1954), pp. 514–22.

AUTHOR

Mary N. Youssef, B.Sc. (Mathematics and Astronomy), 1958, Cairo University, Egypt; M.A. (Mathematical Statistics), 1964, Columbia University; M.Sc.

(Computer Science), 1967, Stanford University; Ph.D. (Statistics), 1970, Oregon State University; AT&T Bell Laboratories, 1970–1983; IBM, 1983—. From 1958 to 1965 Mrs. Youssef was an instructor at Cairo University and the Institute of National Planning, Egypt. At AT&T Bell Laboratories she was engaged in various studies on traffic analysis forecasting and optimal design of switching systems. From 1981 to 1983, she was on leave as Associate Professor of Computer Information Systems at the City University of New York. Currently she is an Advisory Engineer with IBM. Senior Member, IEEE; Member, Sigma Xi, ACM, ORSA.