All drawings appearing in this Recommendation have been done in Autocad.

Recommendation E.507[1)]

MODELS FOR FORECASTING INTERNATIONAL TRAFFIC

1 Introduction

Econometric and time series model development and forecasting requires familiarity with methods and techniques to deal with a range of different situations. Thus, the purpose of this Recommendation is to present some of the basic ideas and leave the explanation of the details to the publications cited in the reference list. As such, this Recommendation is not intended to be a complete guide to econometric and time series modelling and forecasting.

The Recommendation also gives guidelines for building various forecasting models: identification of the model, inclusion of explanatory variables, adjustment for irregularities, estimation of parameters, diagnostic checks, etc.

In addition the Recommendation describes various methods for evaluation of forecasting models and choice of model.

2 Building the forecasting model

This procedure can conveniently be described as four consecutive steps. The first step consists in finding a useful class of models to describe the actual situation. Examples of such classes are simple models, smoothing models, autoregressive models, autoregressive integrated moving average (ARIMA) models or econometric models. Before choosing the class of models, the influence of external variables should be analyzed. If special external variables have significant impact on the traffic demand, one ought to include them in the forecasting models, provided enough historical data are available.

The next step is to identify one tentative model in the class of models which have been chosen. If the class is too extensive to be conveniently fitted directly to data, rough methods for identifying subclasses can be used. Such methods of model identification employ data and knowledge of the system to suggest an appropriate parsimonious subclass of models. The identification procedure may also, in some occasions, be used to yield rough preliminary estimates of the parameters in the model. Then the tentative model is fitted to data by estimating the parameters. Usually, maximum likelihood estimators or least square estimators are used.

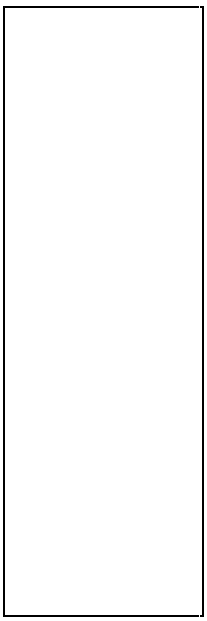
The next step is to check the model. This procedure is often called diagnostic checking. The object is to find out how well the model fits the data and, in case the discrepancy is judged to be too severe, to indicate possible remedies. The outcome of this step may thus be acceptance of the model if the fit is acceptable. If on the other hand it is inadequate, it is an indication that new tentative models may in turn be estimated and subjected to diagnostic checking.

In Figure 1/E.507 the steps in the model building procedure are illustrated.

---

1) The old Recommendation E.506 which appeared in the *Red Book* was split into two Recommendations, revised E.506 and new E.507, and considerable new material was added to both.

Figure 1/E.507 - CCITT 64250



3 Various forecasting models

The objective of § 3 is to give a brief overview of the most important forecasting models. In the GAS 10 Manual on planning data and forecasting methods [5], a more detailed description of the models is given.

3.1 *Curve fitting models*

In curve fitting models the traffic trend is extrapolated by calculating the values of the parameters of some function that is expected to characterize the growth of international traffic over time. The numerical calculations of some curve fitting models can be performed by using the least squares method.

The following are examples of common curve fitting models used for forecasting international traffic:

Linear: $Y_t = a + bt$ (3—1)

Parabolic: $Y_t = a + bt + ct^2$ (3—2)

Exponential: $Y_t = ae^{bt}$ (3—3)

Logistic: $Y_t = \frac{M}{1 + ae^{bt}}$ (3—4)

Gompertz: $Y_t = M(a)^{bt}$ (3—5)

where

Y_t is the traffic at time t,

a, b, c are parameters,

M is a parameter describing the saturation level.

The various trend curves are shown in Figures 2/E.507 and 3/E.507.

The logistic and Gompertz curves differ from the linear, parabolic and exponential curves by having saturation or ceiling level. For further study see [10].

FIGURE 2/E.507 - T0200660-87

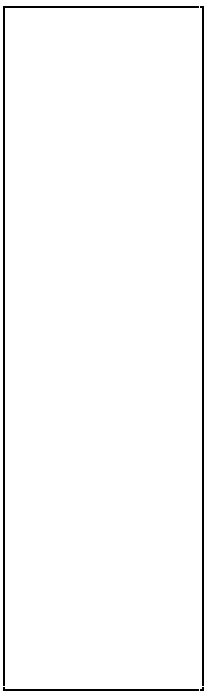
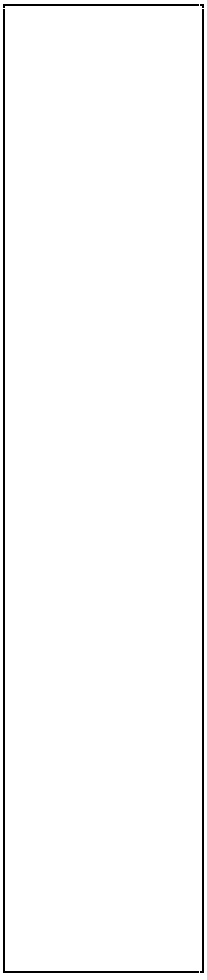


FIGURE 3/E.507 - T0200670-87



3.2 *Smoothing models*

By using a smooth process in curve fitting, it is possible to calculate the parameters of the models to fit current data very well but not necessarily the data obtained from the distant past.

The best known smoothing process is that of the moving average. The degree of smoothing is controlled by the number of most recent observations included in the average. All observations included in the average have the same weight.

In addition to moving average models, there exists another group of smoothing models based on weighting the observations. The most common models are:

— simple exponential smoothing,

— double exponential smoothing,

— discounted regression,

— Holt's method, and

— Holt—Winters' seasonal models.

For example, in the method of exponential smoothing the weight given to previous observations decreases geometrically with age according to the following equation:

$$\hat{\mu}_t = (1 - a)Y_t + a\hat{\mu}_{t-1} \qquad (3—6)$$

where:

Y_t is the measured traffic at time t,

μ_t is the estimated level at time t, and

a is the discount factor [and $(1 - a)$ is the smoothing parameter].

The impact of past observations on the forecasts is controlled by the magnitude of the discount factor.

Use of smoothing models is especially appropriate for short—term forecasts. For further studies see [1], [5] and [9].

3.3 *Autoregressive models*

If the traffic demand, X_t, at time t can be expressed as a linear combination of earlier equidistant observations of the past traffic demand, the process is an autoregressive process. Then the model is defined by the expression:

$$X_t = \Phi_1 X_{t-1} + \Phi_2 X_{t-2} + \ldots + \Phi_p X_{t-p} + a_t \qquad (3—7)$$

where

a_t is white noise at time t;

Φ_k, $k = 1, \ldots p$ are the autoregressive parameters.

The model is denoted by $AR(p)$ since the order of the model is p.

By use of regression analysis the estimates of the parameters can be found. Because of common trends the exogenous variables (X_{t-1}, X_{t-2}, . . . X_{t-p}) are usually strongly correlated. Hence the parameter estimates will be correlated. Furthermore, significance tests of the estimates are somewhat difficult to perform.

Another possibility is to compute the empirical autocorrelation coefficients and then use the Yule—Walker equations to estimate the parameters [Φ_k]. This procedure can be performed when the time series [X_t] are stationary. If, on the other hand, the time series are non stationary, the series can often be transformed to stationarity e.g., by differencing the series. The estimation procedure is given in Annex A, § A.1.

3.4 *Autoregressive integrated moving average (ARIMA) models*

An extention of the class of autoregressive models which include the moving average models is called autoregressive moving average models (ARMA models). A moving average model of order q is given by:

$$X_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} \ldots - \theta_q a_{t-q} \quad (3-8)$$

where

a_t is white noise at time t;

$[\theta_k]$ are the moving average parameters.

Assuming that the white noise term in the autoregressive models in § 3.3 is described by a moving average model, one obtains the so—called ARMA (p, q) model:

$$X_t = \Phi_1 X_{t-1} + \Phi_2 X_{t-2} + \ldots + \Phi_p X_{t-p} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} \ldots - \theta_q a_{t-q} \quad (3-9)$$

The ARMA model describes a stationary time series. If the time series is non—stationary, it is necessary to difference the series. This is done as follow:

Let Y_t be the time series and B the backwards shift operator, then

$$X_t = (1 - B)^d Y_t \quad (3-10)$$

where

d is the number of differences to have stationarity.

The new model ARIMA (p, d, q) is found by inserting equation (3—10) into equation (3—9).

The method for analyzing such time series was developed by G. E. P. Box and G. M. Jenkins [3]. To analyze and forecast such time series it is usually necessary to use a time series program package.

As indicated in Figure 1/E.507 a tentative model is identified. This is carried out by determination of necessary transformations and number of autoregressive and moving average parameters. The identification is based on the structure of the autocorrelations and partial autocorrelations.

The next step as indicated in Figure 1/E.507 is the estimation procedure. The maximum likelihood estimates are used. Unfortunately, it is difficult to find these estimates because of the necessity to solve a nonlinear system of equations. For practical purposes, a computer program is necessary for these calculations. The forecasting model is based on equation (3—9) and the process of making forecasts l time units ahead is shown in § A.2.

The forecasting models described so far are univariate forecasting models. It is also possible to introduce explanatory variables. In this case the system will be described by a transfer function model. The methods for analyzing the time series in a transfer function model are rather similar to the methods described above.

Detailed descriptions of ARIMA models are given in [1], [2], [3], [5], [11], [15] and [17].

3.5 *State space models with Kalman Filtering*

State space models are a way to represent discrete—time process by means of difference equations. The state space modelling approach allows the conversion of any general linear model into a form suitable for recursive estimation and forecasting. A more detailed description of ARIMA state space models can be found in [1].

For a stochastic process such a representation may be of the following form:

$$X_{t+1} = \Phi X_t + Z_t + \omega_t \quad (3-11)$$

and

$$Y_t = HX_t + \nu_t \quad (3-12)$$

where

X_t is an s—vector of state variables in period t,

Z_t is an s—vector of deterministic events,

Φ is an $s \times s$ transition matrix that may, in general, depend on t,

ω_t is an s—vector of random modelling errors,

Y_t is a d—vector of measurements in period t,

H is a $d \times s$ matrix called the observation matrix, and

ν_t is a d—vector of measurement errors.

Both ω_t in equation (3—11) and ν_t in equation (3—12) are additive random sequences with known statistics. The expected value of each sequence is the zero vector and ω_t and ν_t satisfy the conditions:

$$E\left[\omega_t \omega_j^T\right] = Q_t \, \delta_{tj} \text{ for all } t, j, \tag{3—13}$$

$$E\left[\nu_t \nu_j^T\right] = R_t \, \delta_{tj} \text{ for all } t, j,$$

where

Q_t and R_t are nonnegative definite matrices,[2)]

and

δ_{tj} is the Kronecker delta.

Q_t is the covariance matrix of the modelling errors and R_t is the covariance matrix of the measurement errors; the ω_t and the ν_t are assumed to be uncorrelated and are referred to as white noise. In other words:

$$E\left[\nu_t \, \omega_j^T\right] = 0 \text{ for all } t, j, \tag{3—14}$$

and

$$E\left[\nu_t \, X_0^T\right] = 0 \text{ for all } t. \tag{3—15}$$

Under the assumptions formulated above, determine $X_{t,t}$ such that:

$$E\left[(X_{t,t} - X_t)^T (X_{t,t} - X_t)\right] = \text{minimum}, \tag{3—16}$$

where

$X_{t,t}$ is an estimate of the state vector at time t, and

X_t is the vector of true state variables.

The Kalman Filtering technique allows the estimation of state variables recursively for on—line applications. This is done in the following manner. Assuming that there is no explanatory variable Z_t, once a new data point becomes available it is used to update the model:

$$X_{t,t} = X_{t,t-1} + K_t(Y_t - HX_{t,t-1}) \tag{3—17}$$

where

[2)] A matrix A is nonnegative definite, if and only if, for all vectors z, $z^T A z \geq 0$.

K_t is the Kalman Gain matrix that can be computed recursively [18].

Intuitively, the gain matrix determines how much relative weight will be given to the last observed forecast error to correct it. To create a k—step ahead projection the following formula is used:

$$X_{t+k,t} = \Phi^k X_{t,t} \tag{3—18}$$

where

$X_{t+k,t}$ is an estimate of X_{t+k} given observations $Y_1, Y_2, \ldots, Y_t$.

Equations (3—17) and (3—18) show that the Kalman Filtering technique leads to a convenient forecasting procedure that is recursive in nature and provides an unbiased, minimum variance estimate of the discrete time process of interest.

For further studies see [4], [5], [16], [18], [19] and [22].

The Kalman Filtering works well when the data under examination are seasonal. The seasonal traffic load data can be represented by a periodic time series. In this way, a seasonal Kalman Filter can be obtained by superimposing a linear growth model with a seasonal model. For further discussion of seasonal Kalman Filter techniques see [6] and [20].

3.6 *Regression models*

The equations (3—1) and (3—2) are typical regression models. In the equations the traffic, Y_t, is the dependent (or explanatory) variable, while time t is the independent variable.

A regression model describes a linear relation between the dependent and the independent variables. Given certain assumptions ordinary least squares (OLS) can be used to estimate the parameters.

A model with several independent variables is called a multiple regression model. The model is given by:

$$Y_t = ß_0 + ß_1 X_{1t} + ß_2 X_{2t} + \ldots + ß_k X_{kt} + u_t \tag{3—19}$$

where

Y_t is the traffic at time t,

$ß_i$, $i = 0, 1, \ldots, k$ are the parameters,

X_{it}, $i_e = 1, 2, \ldots, k$ is the value of the independent variables at time t,

u_t is the error term at time t.

Independent or explanatory variables which can be used in the regression model are, for instance, tariffs, exports, imports, degree of automation. Other explanatory variables are given in § 2 “Base data for forecasting” in Recommendation E.506.

Detailed descriptions of regression models are given in [1], [5], [7], [15] and [23].

3.7 *Econometric models*

Econometric models involve equations which relate a variable which we wish to forecast (the dependent or endogenous variable) to a number of socio—economic variables (called independent or explanatory variables). The form of the equations should reflect an expected casual relationship between the variables. Given an assumed model form, historical or cross sectional data are used to estimate coefficients in the equation. Assuming the model remains valid over time, estimates of future values of the independent variables can be used to give forecasts of the variables of interest. An example of a typical econometric model is given in Annex C.

There is a wide spectrum of possible models and a number of methods of estimating the coefficients (e.g., least squares, varying parameter methods, nonlinear regression, etc.). In many respects the family of econometric models available is far more flexible than other models. For example, lagged effects can be incorporated, observations weighted, ARIMA residual models subsumed, information from separate sections pooled and parameters allowed to vary in econometric models, to mention a few.

One of the major benefits of building an econometric model to be used in forecasting is that the structure or the process that generates the data must be properly identified and appropriate causal paths must be determined. Explicit structure identification makes the source of errors in the forecast easier to identify in econometric models than in other types of models.

Changes in structures can be detected through the use of econometric models and outliers in the historical data are easily eliminated or their influence properly weighted. Also, changes in the factors affecting the variables in question can easily be incorporated in the forecast generated from an econometric model.

Often, fairly reliable econometric models may be constructed with less observations than that required for time series models. In the case of pooled regression models, just a few observations for several cross—sections are sufficient to support a model used for predictions.

However, care must be taken in estimating the model to satisfy the underlying assumptions of the techniques which are described in many of the reference works listed at the end of this Recommendation. For example the number of independent variables which can be used is limited by the amount of data available to estimate the model. Also, independent variables which are correlated to one another should be avoided. Sometimes correlation between the variables can be avoided by using differenced or detrended data or by transformation of the variables. For further studies see [8], [12], [13], [14] and [21].

4 Discontinuities in traffic growth

4.1 *Examples of discontinuities*

It may be difficult to assess in advance the magnitude of a discontinuity. Often the influence of the factors which cause discontinuties is spread over a transitional period, and the discontinuity is not so obvious. Furthermore, discontinuities arising, for example, from the introduction of international subscriber dialling are difficult to identify accurately, because changes in the method of working are usually associated with other changes (e.g. tariff reductions).

An illustration of the bearing of discontinuities on traffic growth can be observed in the graph of Figure 4/E.507.

Discontinuities representing the doubling — and even more — of traffic flow are known. It may also be noted that changes could occur in the growth trend after discontinuities.

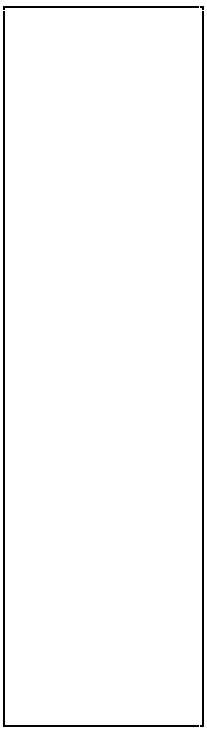
In short—term forecasts it may be desirable to use the trend of the traffic between discontinuities, but for long—term forecasts it may be desirable to use a trend estimate which is based on long—term observations, including previous discontinuities.

In addition to random fluctuations due to unpredictable traffic surges, faults, etc., traffic measurements are also subject to systematic fluctuations, due to daily or weekly traffic flow cycles, influence of time differences, etc.

4.2 *Introduction of explanatory variables*

Identification of explanatory variables for an econometric model is probably the most difficult aspect of econometric model building. The explanatory variables used in an econometric model identify the main sources of influence on the variable one is concerned with. A list of explanatory variables is given in Recommendation E.506, § 2.

Figure 4/E.507 - CCITT 34721



Economic theory is the starting point for variable selection. More specifically, demand theory provides the basic framework for building the general model. However, the description of the structure or the process generating the data often dictate what variables enter the set of explanatory variables. For instance, technological relationships may need to be incorporated in the model in order to appropriately define the structure.

Although there are some criteria used in selecting explanatory variables [e.g., $\overline{R}^2$, Durbin—Watson (D—W) statistic, root mean square error (RMSE), ex—post forecast performance, explained in the references], statistical problems and/or availability of data (either historical or forecasted) limit the set of potential explanatory variables and one often has to revert to proxy variables. Unlike pure statistical models, econometric models admit explanatory variables, not on the basis of statistical criteria alone but, also, on the premise that causality is, indeed, present.

A completely specified econometric model will capture turning points. Discontinuities in the dependent variable will not be present unless the parameters of the model change drastically in a very short time period. Discontinuities in the growth of telephone traffic are indications that the underlying market or technological structure have undergone large changes.

Sustained changes in the growth of telephone demand can either be captured through varying parameter regression or through the introduction of a variable that appears to explain the discontinuity (e.g., the introduction of an advertising variable if advertising is judged to be the cause of the structural change). Once—and—for—all, or step—wise discontinuities, cannot be handled by the introduction of explanatory changes: dummy variables can resolve this problem.

4.3 *Introduction of dummy variables*

In econometric models, qualitative variables are often relevant; to measure the impact of qualitative variables, dummy variables are used. The dummy variable technique uses the value 1 for the presence of the qualitative attribute that has an impact on the dependent variable and 0 for the absence of the given attribute.

Thus, dummy variables are appropriate to use in the case where a discontinuity in the dependent variable has taken place. A dummy variable, for example, would take the value of zero during the historical period when calls were operator handled and one for the period for which direct dial service is available.

Dummy variables are often used to capture seasonal effects in the dependent variable or when one needs to eliminate the effect of an outlier on the parameters of a model, such as a large jump in telephone demand due to a postal strike or a sharp decline due to facility outages associated with severe weather conditions.

Indiscriminate use of dummy variables should be discouraged for two reasons:

1) dummy variables tend to absorb all the explanatory power during discontinuties, and

2) they result in a reduction in the degrees of freedom.

5 Assessing model specification

5.1 *General*

In this section methods for testing the significance of the parameters and also methods for calculating confidence intervals are presented for some of the forecasting models given in § 3. In particular the methods relating to regression analysis and time series analysis will be discussed.

All econometric forecasting models presented here are described as regression models. Also the curve fitting models given in § 3.1 can be described as regression models.

An exponential model given by

$$Z_t = ae^{bt} \cdot u_t \qquad (5—1)$$

may be transformed to a linear form

$$\ln Z_t = \ln a + bt + \ln u_t \qquad (5—2)$$

or

$$Y_t = ß_0 + ß_1 X_t + a_t \qquad (5—3)$$

where

Y_t $= \ln Z_t$

$ß_0$ $= \ln a$

$ß_1$ $= b$

X_t $= t$

a_t $= \ln u_t$ (white noise).

5.2 *Autocorrelation*

A good forecasting model should lead to small autocorrelated residuals. If the residuals are significantly correlated, the estimated parameters and also the forecasts may be poor. To check whether the errors are correlated, the autocorrelation function r_k, $k = 1, 2, \ldots$ is calculated. r_k is the estimated autocorrelation of residuals at lag k. A way to detect autocorrelation among the residuals is to plot the autocorrelation function and to perform a Durbin—Watson test. The Durbin—Watson statistic is:

$$\text{D—W} = \frac{\sum_{t=2}^{N} (e_t - e_{t-1})^2}{\sum_{t=1}^{N} et^2} \qquad (5—4)$$

where

e_t is the estimated residual at time t,

N is the number of observations.

5.3 *Test of significance of the parameters*

One way to evaluate the forecasting model is to analyse the impact of different exogenous variables. After estimating the parameters in the regression model, the significance of the parameters has to be tested.

In the example of an econometric model in Annex C, the estimated values of the parameters are given. Below these values the estimated standard deviation is given in parentheses. As a rule of thumb, the parameters are considered as significant if the absolute value of the estimates exceeds twice the estimated standard deviation. A more accurate way of testing the significance of the parameters is to take into account the distributions of their estimators.

The multiple correlation coefficient (or coefficient of determination) may be used as a criterion for the fitting of the equation.

The multiple correlation coefficient, R^2, is given by:

$$R^2 = \frac{\sum_{i=1}^{N} (\hat{Y}_j - \overline{Y})^2}{\sum_{i=1}^{N} (Y_i - \overline{Y})^2} \qquad (5—5)$$

If the multiple correlation coefficient is close to 1 the fitting is satisfactory. However, a high R^2 does not imply an accurate forecast.

In time series analysis, the discussion of the model is carried out in another way. As pointed out in § 3.4, the number of autoregressive and moving average parameters in an ARIMA model is determined by an identification procedure based on the structure of the autocorrelation and partial autocorrelation function.

The estimation of the parameters and their standard deviations is performed by an iterative nonlinear estimation procedure. Hence, by using a time series analysis computer program, the estimates of the parameters can be evaluated by studying the estimated standard deviations in the same way as in regression analysis.

An overall test of the fitting is based on the statistic

$$Q_{N-d} = \sum_{i=1}^{N} r_i^2 \qquad (5—6)$$

where r_i is the estimated autocorrelation at lag i and d is the number of parameters in the model. When the model is adequate, Q_{N-d} is approximately chi—square distributed with $N - d$ degrees of freedom. To test the fitting, the value Q_{N-d} can be compared with fractiles of the chi—square distribution.

5.4 *Validity of exogenous variables*

Econometric forecasting models are based on a set of exogenous variables which explain the development of the endogenous variable (the traffic demand). To make forecasts of the traffic demand, it is necessary to make forecasts of each of the exogenous variables. It is very important to point out that an exogenous variable should not be included in the forecasting model if the prediction of the variable is less confident than the prediction of the traffic demand.

Suppose that the exact development of the exogenous variable is known which, for example, is the case for the simple models where time is the explanatory variables. If the model fitting is good and the white noise is normally distributed with expectation equal to zero, it is possible to calculate confidence limits for the forecasts. This is easily done by a computer program.

On the other hand, the values of most of the explanatory variables cannot be predicted exactly. The confidence of the prediction will then decrease with the number of periods. Hence, the explanatory variables will cause the confidence interval of the forecasts to increase with the number of the forecast periods. In these situations it is difficult to calculate a confidence interval around the forecasted values.

If the traffic demand can be described by an autoregressive moving average model, no explanatory variables are included in the model. Hence, if there are no explanatory variable in the model, the confidence limits of the forecasting values can be calculated. This is done by a time series analysis program package.

5.5 *Confidence intervals*

Confidence intervals, in the context of forecasts, refer to statistical constructs of forecast bounds or limits of prediction. Because statistical models have errors associated with them, parameter estimates have some variability associated with their values. In other words, even if one has identified the correct forecasting model, the influence of endogenous factors will cause errors in the parameter estimates and the forecast. Confidence intervals take into account the uncertainty associated with the parameter estimates.

In causal models, another source of uncertainty in the forecast of the series under study are the predictions of the explanatory variables. This type of uncertainty cannot be handled by confidence intervals and is usually ignored, even though it may be more significant than the uncertainty associated with coefficient estimates. Also, uncertainty due to possible outside shocks is not reflected in the confidence intervals.

For a linear, static regression model, the confidence interval of the forecast depends on the reliability of the regression coefficients, the size of the residual variance, and the values of the explanatory variables. The 95% confidence interval for a forecasted value Y_{N+1} is given by:

$$\hat{Y}_N(1) — 2\hat{\sigma}\ Y_{N+1}\ \hat{Y}_N(1) + 2\hat{\sigma} \qquad (5—7)$$

where $\hat{Y}_N(1)$ is the forecast one step ahead and $\sigma\not{c}$ is the standard error of the forecast.

This says that we expect, with a 95% probability, that the actual value of the series at time $N + 1$ will fall within the limits given by the confidence interval, assuming that there are no errors associated with the forecast of the explanatory variables.

6 Comparison of alternative forecasting models

6.1 *Diagnostic check — Model evaluation*

Tests and diagnostic checks are important elements in the model building procedure. The quality of the model is characterized by the residuals. Good forecasting models should lead to small autocorrelated residuals, the variance of the residuals should not decrease or increase and the expectation of the residuals should be zero or close to zero. The precision of the forecast is affected by the size of the residuals which should be small.

In addition the confidence limits of the parameter estimates and the forecasts should be relatively small. And in the same way, the mean square error should be small compared with results from other models.

6.2 *Forecasts of levels versus forecasts of changes*

Many econometric models are estimated using levels of the dependent and independent variables. Since economic variables move together over time, high coefficients of determination are obtained. The collinearity among the levels of the explanatory variables does not present a problem when a model is used for forecasting purposes alone, given that the collinearity pattern in the past continues to exist in the future. However, when one attempts to measure structural coefficients (e.g., price and income elasticities) the collinearity of the explanatory variables (known as multicollinearity) renders the results of the estimated coefficients unreliable.

To avoid the multicollinearity problem and generate benchmark coefficient estimates and forecasts, one may use changes of the variables (first difference or first log difference which is equivalent to a percent change) to estimate a model and forecast from that model. Using changes of variables to estimate a model tends to remove the effect of multicollinearity and produce more reliable coefficient estimates by removing the common effect of economic influences on the explanatory variables.

By generating forecasts through levels of and changes in the explanatory variables, one may be able to produce a better forecast through a reconciliation process. That is, the models are adjusted so that the two sets of forecasts give equivalent results.

6.3 *Ex—post forecasting*

Ex—post forecasting is the generation of a forecast from a model estimated over a sub—sample of the data beginning with the first observation and ending several periods prior to the last observation. In ex—post forecasting, actual values of the explanatory variables are used to generate the forecast. Also, if forecasted values of the explanatory variables are used to produce an ex—post forecast, one can then measure the error associated with incorrectly forecasted explanatory variables.

The purpose of ex—post forecasting is to evaluate the forecasting performance of the model by comparing the forecasted values with the actuals of the period after the end of the sub—sample to the last observation. With ex—post forecasting, one is able to assess forecast accuracy in terms of:

1) percent deviations of forecasted values from actual values,

2) turning point performance,

3) systematic behaviour of deviations.

Deviations of forecasted values from actual values give a general idea of the accuracy of the model. Systematic drifts in deviations may provide information for either re—specifying the model or adjusting the forecast to account for the drift in deviations. Of equal importance in evaluating forecast accuracy is turning point performance, that is, how well the model is able to forecast changes in the movement of the dependent variable. More criteria for evaluating forecast accuracy are discussed below.

6.4 *Forecast performance criteria*

A model might fit the historical data very well. However, when the forecasts are compared with future data that are not used for estimation of parameters, the fit might not be so good. Hence comparison of forecasts with actual observations may give additional information about the quality of the model. Suppose we have the time series, Y_1, Y_2, ..., Y_N, Y_{N+1}, ..., Y_{N+M}.

The M last observations are removed from the time series and the model building procedure. The one—step—ahead forecasting error is given by:

$$e_{N+t} = Y_{N+t} - \hat{Y}_{N+t-1}(1) \quad t = 1, 2, \ldots, M \tag{6—1}$$

where

$\hat{Y}_{N+t-1}(1)$ is the one—step—ahead forecast.

Mean error

The mean error, ME, is defined by

$$\mathrm{ME} = \frac{1}{M} \sum_{t=1}^{M} e_{N+t} \tag{6—2}$$

ME is a criterium for forecast bias. Since the expectation of the residuals should be zero, a large deviation from zero indicates bias in the forecasts.

Mean percent error

The mean percent error, MPE, is defined by

$$\mathrm{MPE} = \frac{100}{M} \sum_{t=1}^{M} \frac{e_{n+t}}{Y_{N+t}} \tag{6—3}$$

This statistic also indicates possible bias in the forecasts. The criterium measures percentage deviation in the bias. It is not recommended to use MPE when the observations are small.

Root mean square error

The root mean square error, RMSE, of the forecast is defined as

$$\text{RMSE} = \left[\frac{1}{M} \sum_{t=1}^{M} e_{N+t}^{2} \right]^{1/2} \qquad (6—4)$$

RMSE is the most commonly used measure for forecasting precision.

Mean absolute error

The mean absolute error, MAE, is given by

$$\text{MAE} = \frac{1}{M} \sum_{t=1}^{M} \left| e_{N+t} \right| \qquad (6—5)$$

Theil's inequality coefficient

Theil's inequality coefficient is defined as follows:

$$U = \left[\sum_{t=1}^{M} \frac{e_{N+t}^{2}}{Y_{N+t}^{2}} \right]^{1/2} \qquad (6—6)$$

Theil's U is preferred as a measure of forecast accuracy because the error between forecasted and actual values can be broken down to errors due to:

1) central tendency,

2) unequal variation between predicted and realized changes, and

3) incomplete covariation of predicted and actual changes.

This decomposition of prediction errors can be used to adjust the model so that the accuracy of the model can be improved.

Another quality that a forecasting model must possess is ability to capture turning points. That is, a forecast must be able to change direction in the same time period that the actual series under study changes direction. If a model is estimated over a long period of time which contains several turning points, ex—post forecast analysis can generally detect a model's inability to trace closely actuals that display turning points.

7 Choice of forecasting model

7.1 *Forecasting performance*

Although the choice of a forecasting model is usually guided by its forecasting performance, other considerations must receive attention. Thus, the length of the forecast period, the functional form, and the forecast accuracy of the explanatory variables of an econometric model must be considered.

The length of the forecast period affects the decision to use one type of a model versus another, along with historical data limitations and the purpose of the forecasting model. For instance, ARIMA models may be appropriate forecasting models for short—term forecasts when stability is not an issue, when sufficient historical data are available, and when causality is not of interest. Also, when the structure that generates the data is difficult to identify, one has no choice but to use a forecasting model which is based on historical data of the variable of interest.

The functional form of the model must also be considered in a forecasting model. While it is true that a more complex model may reduce the model specification error, it is also true that it will, in general, considerably increase the effect of data errors. The model form should be chosen to recognize the trade—off between these sources of error.

Availability of forecasts for explanatory variables and their reliability record is another issue affecting the choice of a forecasting model. A superior model using explanatory variables which may not be forecasted accurately can be inferior to an average model whose explanatory variables are forecasted accurately.

When market stability is an issue, econometric models which can handle structural changes should be used to forecast. When causality matters, simple models or ARIMA models cannot be used as forecasting tools. Nor can they be used when insufficient historical data exist. Finally, when the purpose of the model is to forecast the effects associated with changes in the factors that influence the variable in question, time series models may not be appropriate (with the exception, of course, of transfer function and multiple time series models).

7.2 *Length of forecast period*

For normal extensions of switching equipment and additions of circuits, a forecast period of about six years is necessary. However, a longer forecast period may be necessary for the planning of new cables or other transmission media or for major plant installations. Estimates in the long term would necessarily be less accurate than short—term forecasts but that would be acceptable.

In forecasting with a statistical model, the length of the forecast period is entirely determined by:

a) the historical data available,

b) the purpose or use of the forecast,

c) the market structure that generates the data,

d) the forecasting model used,

e) the frequency of the data.

The historical data available depends upon the period over which it has been collected and the frequency of collection (or the length of the period over which data is aggregated). A small historical data base can only support a short prediction interval. For example, with 10 or 20 observations a model can be used to forecast 4—5 periods past the sample (i.e. into the future). On the other hand, with 150—200 observations, potentially reliable forecasts can be obtained for 30 to 50 periods past the sample — other things being equal.

Certainly, the purpose of the forecast affects the number of predicted periods. Long range facility planning requires forecasts extending 15—20 or more years into the future. Rate change evaluations may only require forecasts for 2—3 years. Alteration of routing arrangements could only require forecasts extending a few months past the sample.

Stability of a market, or lack thereof, also affect the length of the forecast period. With a stable market structure one could conceivably extend the forecast period to equal the historical period. However, a volatile market does not afford the same luxury to the forecaster; the forecast period can only consist of a few periods into the future.

The forecasting models used to generate forecasts do, by their nature, influence the decision on how far into the future one can reasonably forecast. Structural models tend to perform better than other models in the long run, while for short—run predictions all models seem to perform equally well.

It should be noted that while the purpose of the forecast and the forecasting model affect the length of the forecast, the number of periods to be forecasted play a crucial role in the choice of the forecasting model and the use to which a forecast is put.

ANNEX A

(to Recommendation E.507)

DESCRIPTION OF FORECASTING PROCEDURES

A.1 *Estimation of autoregressive parameters*

The empirical autocorrelation at lag k is given by:

$$r_k = \frac{v_k}{v_0} \qquad \text{(A—1)}$$

where

$$v_k = \frac{1}{N-1} N-kt = 1 \ (X_t - \overline{X}) \ (X_{t+k} - \overline{X}) \tag{A—2}$$

and

$$\overline{X} = \frac{1}{N} \sum_{t=1}^{N} X_t \tag{A—3}$$

N being the total number of observations.

The relation between $[r_k]$ and the estimates $[\hat{\Phi}_k]$ of $[\Phi_k]$ is given by the Yule—Walker equations:

$$\begin{aligned} r_1 &= \hat{\Phi}_1 + \hat{\Phi}_2 r_1 + \ldots + \hat{\Phi}_p r_{p-1} \\ r_2 &= \hat{\Phi}_1 r_1 + \hat{\Phi}_2 r_2 \ldots \hat{\Phi}_p r_{p-2} \\ &\cdot \\ &\cdot \\ &\cdot \\ r_p &= \hat{\Phi}_1 r_{p-1+} \hat{\Phi}_2 r_{p-2} + \ldots + \hat{\Phi}_p \end{aligned} \tag{A—4}$$

Hence the estimators $[\hat{\Phi}_k]$ can be found by solving this system of equations.

For computations, an alternative to directly solving the equations is the following recursive procedure. Let $[\hat{\Phi}_{k,\,j}]_j$ be estimators of the parameters at lag j = 1, 2, . . ., k given that the total number of parameters are k. The estimators $[\hat{\Phi}_{k+1,\,j}]_j$ are then found by

$$\hat{\Phi}_{k+1,\,k+1} = \frac{r_{k+1} \sum_{j=1}^{k} \hat{\Phi}_{k;j} \, r_{k+1-j}}{1 - \sum_{j=1}^{k} \hat{\Phi}_{k;j} \, r_j} \tag{A—5}$$

$$\hat{\Phi}_{k+1,\,j} = \hat{\Phi}_{kj} - \hat{\Phi}_{k+1,\,k+1} \, \hat{\Phi}_{k,k-j+1} \qquad j = 1, 2, \ldots, k \tag{A—6}$$

Defining $\hat{\Phi}_{p,\,j} = \hat{\Phi}_j$, j = 1, 2, . . ., p, the forecast of the traffic demand at time t+1 is expressed by:

$$X_{t+1} = \hat{\Phi}_1 X_t + \hat{\Phi}_2 X_{t-1} + \ldots + \hat{\Phi}_p X_{t-p} \tag{A—7}$$

A.2 *Forecasting with ARIMA models*

The forecast l time units ahead is given by:

$$\begin{aligned}\hat{X}_t(l) = \hat{\Phi}_1\,[X_{t+l-1}] + \hat{\Phi}_2\,[X_{t+l-2}] \\ + \ldots + \hat{\Phi}_p[X_{t+l-p}] \\ + [a_{t+l}] - \hat{\theta}_1\,[a_{t+l-1}] \\ - \hat{\theta}_2[a_{t+l-2}] - \ldots - \hat{\theta}_q[a_{t+l-q}]\end{aligned} \qquad \text{(A—8)}$$

where $$[\hat{X}_j] = \begin{cases} \hat{X}_t(j-t) & \text{if } j > t \\ X_j & \text{if } j \le t \end{cases} \qquad \text{(A—9)}$$

$$[a_j] = \begin{cases} 0 & \text{if } j > t \\ X_j - \hat{X}_j & \text{if } j \le t \end{cases} \qquad \text{(A—10)}$$

which means that $[X_j]$ is defined as a forecast when $j > t$ and otherwise as an actual observation and that $[a_j]$ is defined as 0 when $j > t$ since white noise has expectation 0. If the observations are known ($j \le t$), then $[a_j]$ is equal to the residual.

ANNEX B

(to Recommendation E.507)

KALMAN FILTER FOR A LINEAR TREND MODEL

To model telephone traffic, it is assumed that there are no deterministic changes in the demand pattern. This situation can be modelled by setting the deterministic component Z_t to zero. Then the general state space model is:

$$X_{t+1} = X_t + \omega_t \qquad \text{(B—1)}$$

$$Y_t = HX_t + \nu_t$$

where

X_t is an s—vector of state variables in period t,

Y_t is a vector of measurements in year t,

φ is an $s \times s$ transition matrix that may, in general, depend on t,

and

ω_t is an s—vector of random modelling errors,

ν_t is the measurement error in year t.

For modelling telephone traffic demand, adapt a simple two—state, one—data variable model defined by:

$$X_{t+1} = \begin{bmatrix} x_{t+1} \\ \dot{x}_{t+1} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x_t \\ \dot{x}_t \end{bmatrix} + \begin{bmatrix} \omega_t \\ \dot{\omega}_t \end{bmatrix} \qquad \text{(B—2)}$$

and

$$y_t = x_t + \nu_t \qquad \text{(B—3)}$$

where

x_t is the true load in year t,

$\dot{x}_t$ is the true incremental growth in year t,

y_t is the measured load in year t,

ν_t is the measurement error in year t.

Thus, in our model

$$\varphi = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \quad \text{and } H = 1. \tag{B—4}$$

The one—step—ahead projection is written as follows:

$$X_{t+1,t} = \begin{bmatrix} x_{t+1.t} \\ \dot{x}_{1.t} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \quad \begin{bmatrix} x_{t.t} \\ \dot{x}_{t.t} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \quad \begin{bmatrix} x_{t,t-1} + \alpha_t(y_t - x_{t,t-1}) \\ \dot{x}_{t.t-1} + ß_t(y_t - x_{t,t-1}) \end{bmatrix} \tag{B—5}$$

where

$X_{t+1,t}$ is the projection of the state variable in period $t + 1$ given observations through year t.

The α_t and $ß_t$ coefficients are the Kalman gain matrices in year t. Rewriting the above equation yields:

$$x_{t,t} = (1-\alpha_t)x_{t,t-1} + \alpha_t y_t \tag{B—6}$$

and

$$\dot{x}_{t,t} = (1-ß_t)_t \dot{x}_{t1} + ß_t(y_t - x_{t-1,t-1}) \tag{B—7}$$

The Kalman Filter creates a linear trend for each time series being forecast based on the current observation or measurement of traffic demand and the previous year's forecast of that demand. The observation and forecasted traffic load are combined to produce a smoothed load that corresponds to the level of the process, and a smoothed growth increment. The Kalman gain values α_t and $ß_t$ can be either fixed or adaptive. In [16] Moreland presents a method for selecting fixed, robust parameters that provide adequate performance independent of system noise, measurement error, and initial conditions. For further details on the proper selection of these parameters see [6], [20] and [22].

ANNEX C

(to Recommendation E.507)

EXAMPLE OF AN ECONOMETRIC MODEL

To illustrate the workings of an econometric model, we have chosen the model of United States billed minutes to Brazil. This model was selected among alternative models for three reasons:

a) to demonstrate the introduction of explanatory variables,

b) to point out difficulties associated with models used for both the estimation of the structure and forecasting purposes, and

c) to show how transformations may affect the results.

The demand of United States billed minutes to Brazil (*MIN*) is estimated by a log—linear equation which includes United States billed messages to Brazil (*MSG*), a real telephone price index (*RPI*), United States personal income in 1972 prices (*YP*72), and real bilateral trade between the United States and Brazil (*RTR*) as explanatory variables. This model is represented as:

$$\ln(MIN)_t = ß_0 + ß_1 \ln(MSG)_t + ß_2 \ln(RPI)_t + ß_3 \ln(YP72)_t + ß_4 \ln(RTR)_t + u_t \quad \text{(C—1)}$$

where u_t is the error term of the regression and where, $ß_1 > 0$, $ß_2 < 0$, $ß_3 > 0$ and $ß_4 > 0$ are expected values.

Using ridge regression to deal with severe multicollinearity problems, we estimate the equation over the 1971 : 1 (i.e. first quarter of 1971) to 1979 : 4 interval and obtain the following results:

$$\ln(MIN)_t = —3.489 + (0.619) \ln(MSG)_t — (0.447) \ln(RPI)_t + (1.166) \ln(YP72)_t + (0.281) \ln(RTR)_t$$

$$\text{In}(MIN)_t = —3.489 + (0.035) \ln(MSG)_t — (0.095) \ln(RPI)_t + (0.269) \ln(YP72)_t + (0.084) \quad \text{(C—2)}$$

$$\overline{R}^2 = 0.985,\ SER = 0.083,\ D—W = 0.922,\ k = 0.10 \quad \text{(C—3)}$$

where $\overline{R}^2$ is the adjusted coefficient of determination, *SER* is the standard error of the regression, D—W is the Durbin—Watson statistic, and k is the ridge regression constant. The values in parentheses under the equation are the estimated standard deviation of the estimated parameters $\hat{ß}_1$, $\hat{ß}_2$, $\hat{ß}_3$, $\hat{ß}_4$.

The introduction of messages as an explanatory variable in this model was necessitated by the fact that since the mid—seventies transmission quality has improved and completion rates have risen while, at the same time, the strong growth in this market has begun to dissipate. Also, the growth rates for some periods could not have been explained by rate activity on either side or real United States personal income. The behaviour of the message variable in the minute equation was able to account for all these factors.

Because the model serves a dual purpose — namely, structure estimation and forecasting — at least one more variable is introduced than if the model were to be used for forecasting purposes alone. The introduction of additional explanatory variables results in severe multicollinearity and necessitates employing ridge regression which lowers $\overline{R}^2$ and the Durbin—Watson statistic. Consequently, the predictive power of the model is reduced somewhat.

The effect of transforming the variables of a model are shown in the ex—post forecast analysis performed on the model of United States billed minutes to Brazil. The deviations using levels of the variables are larger than those of the logarithms of the variables which were used to obtain a better fit (the estimated RMSE for the log—linear regression model is 0.119 827). The forecast results in level and logarithmic form are shown in Table C—1/E.507.

TABLE C—1/E.507

	Logarithms			Levels		
	Forecast	Actual	% deviation	Forecast	Actual	% deviation
1980: 1	14.858	14.938	—0.540	2 836 269	3 073 697	— 7.725
2	14.842	14.972	—0.872	2 791 250	3 180 334	—12.234
3	14.916	15.111	—1.296	3 005 637	3 654 092	—17.746
4	14.959	15.077	—0.778	3 137 698	3 529 016	—11.089
1981: 1	15.022	15.102	—0.535	3 341 733	3 621 735	— 7.731
2	14.971	15.141	—1.123	3 175 577	3 762 592	—15.601
3	15.395	15.261	—0.879	4 852 478	4 244 178	14.333
4	15.405	15.302	—0.674	4 901 246	4 421 755	—10.844
1982: 1	15.365	15.348	—0.110	4 709 065	4 630 238	— 1.702
2	15.326	15.386	—0.387	4 528 947	4 807 901	— 5.802

References

[1] ABRAHAM (A.) and LEDOLTER (J.): Statistical methods for forecasting *J. Wiley*, New York, 1983.

[2] ANDERSON (O. D.): Time series analysis and forecasting. The Box—Jenkins approach. *Butterworth*, London, 1976.

[3] BOX (G. E. P.) and JENKINS (G. M.): Time Series Analysis: Forecasting and Control, *Holden—Day*, San Francisco, 1976.

[4] BROWN (R. G.): Introduction to random signal analysis and Kalman Filtering. *John Wiley & Sons*, New York, 1983.

[5] CCITT: Manual planning data and forecasting methods, Vol. I and II, *ITU*, Geneva, 1988.

[6] CHEMOUIL (P.) and GARNIER (B.): An Adaptive Short—Term Traffic Forecasting Procedure Using Kalman Filtering. *ITC 11*, Tokyo, 1985.

[7] DRAPER (N.) and SMITH (H.): Applied Regression Analysis, Second Edition, *John Wiley & Sons*, New York, 1981.

[8] DUTTA (M.): Econometric Methods, *South—Western Publishing Co.*, Cincinnati, 1975.

[9] GARDNER (E. S. Jr.): Exponential smoothing the state of art. *Journal of forecasting*, 4, pp. 1—28, 1985.

[10] GILCHRIST W.: Statistical forecasting. *John Wiley & Sons*, New York, 1976.

[11] GRANGER (C. W. J.) and NEWBOLD (P.): Forecasting Economic Time Series, *Academic Press*, New York, 1977.

[12] JOHNSTON (J.): Econometric Methods, Second Edition, *McGraw—Hill*, New York, 1972.

[13] JUDGE (G. G.) *et al.*: The Theory and Practice of Econometrics, *John Wiley & Sons*, New York, 1980.

[14] KMENTA (J.): Elements of Econometrics, *Macmillan Publishing Company*, New York, 1971.

[15] MAKRIDAKIS (S.), WHEELWRIGHT (S. C.), McGEE (V. .E.): Forecasting methods and applications Second Edition. *John Wiley & Sons*, New York, 1983.

[16] MORELAND (J. P.): A robust sequential projection algorithm for traffic load forecasting. *The Bell Technical Journal*, Vol. 61, No. 1, 1982.

[17] NELSON (C. R.): Applied Time Series Analysis for Managerial Forecasting, *Holden—Day*, San Francisco, 1973.

[18] PACK (C. D.) and WHITAKER (B. A.): Kalman Filter models for network forecasting. *The Bell Technical Journal*, Vol. 61, No. 1, pp. 1—9, 1982.

[19] SORENSON (H. W.): Kalman filtering techniques. Advances in control systems theory and applications. *Academic Press*, Vol. 3, pp. 219—292, 1966.

[20] SZELAG (C. R.): A short—term forecasting algorithm for trunk demand servicing. *The Bell Technical Journal*, Vol. 61, No. 1, pp. 67—96, 1982.

[21] THEIL (H.): Principles of Econometrics, *John Wiley & Sons*, New York, 1971.

[22] TOME (F. M.) and CUNHA (J. A.): Traffic forecasting with a state space model. *ITC 11*, Tokyo, 1985.

[23] WONNACOTT (T. H.) and WONNACOTT (R. J.): Regression. *John Wiley & Sons*, New York, 1981.

Bibliography

PINDYCK (R. S.) and RUBINFELD (D. F.): Econometric Models and Econometric Forecasts, *McGraw—Hill*, New York, 1981.

SASTRI, (T.): A state space modelling approach for time series forecasting. *Management Science*, Vol. 31, No. 11, pp. 1451—1470, 1985.