Forecast Pro
Statistical Reference Manual

Eric A. Stellwagen and Robert L. Goodrich
Business Forecast Systems, Inc.

Manual Last Revised: December 5, 2017

Business Forecast Systems, Inc.
68 Leonard Street, Belmont, MA 02478 USA
Phone: 617-484-5050 ◆ Fax: 617-484-9219
Email: info@forecastpro.com ◆ Web: www.forecastpro.com

Contents

Statistical Reference

Expert Selection **2**

Simple Methods **2**

Exponential Smoothing **3**
Conceptual Overview 4
Models of the Exponential Smoothing Family 5
Implementation of Exponential Smoothing in Forecast Pro 7
Statistical Description of Exponential Smoothing 9

Discrete Distributions **16**
Poisson Distribution 16
Negative Binomial Distribution 17

Croston's Intermittent Demand Model **18**

Curve Fitting **19**

Box-Jenkins Statistical Models **20**
Implementation of Box-Jenkins in Forecast Pro 20
Box-Jenkins Background 21
Description of the ARIMA Model 23
Seasonal Models 26
Selecting Model Orders 27

Dynamic Regression **28**
Description of Dynamic Regression Model 29
Dynamic Regression Diagnostics 30

Bass Diffusion Model **32**

Forecasting By Analogy **33**

Model Statistics **34**

Box-Cox Power Transforms **36**

Safety Stocks **37**

Outlier Detection and Correction **39**

Trading Day Effects **39**

Methodology of Automatic Forecasting **41**

Introduction 41
Classification of Time Series 42
Multiple-level Forecasting 43
Incorporation of Additional Information 44
Selection of Forecasting Method 44
Model Selection via Out-of-Sample Testing 46

Glossary **49**

Bibliography **53**

Index **59**

Statistical Reference

This manual describes the statistical techniques, statistics, and strategies that are implemented in Forecast Pro. It is not necessary that you fully understand, or even read, this manual in order to produce accurate forecasts with the product.

Those who would like a more thorough coverage of this topic should consult the book *Applied Statistical Forecasting* or any of the other texts found in the bibliography. *Applied Statistical Forecasting* was written by Dr. Robert L. Goodrich, the author of Forecast Pro, and is available from Business Forecast Systems.

This chapter begins by presenting each of the forecasting models and concludes with a discussion of the model statistics presented by the program. The topics are:

Expert selection

Simple methods

Exponential smoothing

Discrete distributions

Croston’s intermittent demand model

Curve fitting

Box-Jenkins

Dynamic regression

Model statistics

Expert Selection

Expert selection allows Forecast Pro to select an appropriate univariate forecasting technique automatically. Expert selection operates as follows.

If the data set is very short, Forecast Pro defaults to simple moving average.

Otherwise Forecast Pro examines the data for the applicability of the intermittent or discrete forecast models. Although the forecasts produced from such models are just straight horizontal lines, they often provide forecasts superior to those from exponential smoothing for low-volume, messy data.

If neither of these models are applicable to the data, the choice is now narrowed down to different forms of exponential smoothing and Box-Jenkins models. Forecast Pro next runs a series of tests on the data and applies a rule-based logic that may lead to a model selections based on data characteristics.

If the rule-based logic does not lead to a definitive answer, Forecast Pro performs an out-of-sample test to choose between an exponential smoothing model and a Box-Jenkins model.

There is also a question of selecting the form of the exponential smoothing and Box-Jenkins models. This procedure is documented in the *Implmentation of Exponential Smoothing in Forecast Pro* and *Implmentation of Box-Jenkins in Forecast Pro* sections of this manual.

Simple Methods

Forecast Pro supports three variants of the n-term simple moving average, which we symbolize as SMA(n). The essence SMA(n) is to estimate the current *level* S_t of the series as the average of the last n observations. The level of the series is defined as the value that the observation would take if it were not obscured by noise.

$$S_t = \frac{1}{n}\sum_{s=0}^{n-1} Y_{t-s}$$

The forecast for time t+m from the forecast base t is simply a horizontal line at the level S_t.

$$\hat{Y}_t(m) = S_t$$

Confidence limits for SMA(n) are determined by assuming that the true underlying process is a random walk with observation error.

SMA(n) has one purpose—to decrease the effect of noise on the estimated true value of the series. It cannot pick up the effects of seasonality or trending. Thus its capabilities are very similar to those of simple exponential smoothing, except that the model has no parameters that need to be fitted to the data.

SMA(n) should be used only when the historical data record is so short and so noisy that it is meaningless to try to extract patterns from the data or even to estimate a smoothing weight. In any other circumstance, one of the exponential smoothing models will outperform SMA(n).

Forecast Pro offers three versions of SMA(n)—Automatic, Moving average and Random walk. Automatic determines the number of terms n in the moving average by determining the n that minimizes error over the historic sample. Moving average lets the user set n. Random walk sets n to 1, so that the forecast consists of the last observed value.

Exponential Smoothing

Exponential smoothing is the most widely applicable of the univariate time series methods to business data. In the absence of information to the contrary, it is probably the best choice for the typical user.

Although exponential smoothing was first developed over thirty years ago, it is still very much a hot topic in research circles. If anything, its reputation as a robust, easy to understand methodology has increased in recent years, often at the expense of Box-Jenkins.

The main reason for this is that Box-Jenkins models are built upon the abstract statistical concept of autocorrelation, while exponential smoothing models are built upon clear-cut features like level, trend, and seasonality. Exponential smoothing models are therefore less likely to be influenced by purely statistical quirks in the data.

Harvey [1984, 1990] has extended the exponential smoothing approach in his development of so-called *structural models*. Structural model forecasts are generated from a Kalman filter built upon a formal statistical model involving the same features as exponential smoothing—level, trend and seasonality. We now recognize exponential smoothing for what it really is—approximate Kalman filters fitted directly to the data.

This establishes a framework for extending the basic exponential smoothing methodology. You will see two such extensions in the methodological descriptions below.

> *Proportional error* models extend exponential smoothing to the case where errors tend to be proportional to the level of the data. The majority of business data seem to exhibit this trait.
>
> *Event adjustment* models extend exponential smoothing to include the estimation of, and adjustment for, promotional or other nonperiodic events.

Conceptual Overview

Exponential smoothing is based upon a structural model of time series data. We assume that the time series process manifests some or all of the following structural components.

> **Level**. The level of a time series is a smooth, slowly changing, nonseasonal process underlying the observations. We cannot measure the level directly because it is obscured by seasonality, promotional events and irregularity (noise). It must be estimated from the data.
>
> **Local Trend**. The local trend is the smooth, slowly changing rate of change of the level. We call it *local* to emphasize the fact that at each point in time it undergoes a small but unpredictable change. Forecasts are based on the local trend at the end of the historic data, not the overall global trend. We cannot measure the trend directly. It must be estimated from the data.
>
> **Seasonal Effects**. Additive or multiplicative seasonal indexes represent periodic patterns in the time series, like the annual patterns in retail sales. Like the level and the trend, seasonal indexes must be estimated from the data. They are assumed to undergo small changes at each point in time.
>
> **Event Effects**. Promotional events influence sales in much the same way as seasonality but they are not usually periodic. Additive or multiplicative event indexes are estimated from the data in much the same way as seasonal indexes. They are assumed to undergo small changes at each point in time.
>
> **Random Events**. The level, local trend, seasonal and event indexes are all stochastic—that is their values change unpredictably from point to point in time. These changes are caused by unpredictable events like the amount by which a company's actual profit or loss differs from what was expected. These are often called *random shocks*.
>
> **Noise**. All of the features described so far are components of an ongoing historical process. Our measurements of the process, however, are usually corrupted by noise or measurement error. For instance, chewing gum shipments or chewing gum orders are noisy measurements of chewing gum consumption.

Three of these features—level, random events and noise—are present in every exponential smoothing model. The remaining three—local trend, seasonal indexes and event effects—may be present or absent. We *identify* a model by determining which of these features should be included to describe the data properly.

Originally, exponential smoothing models were built informally on these features, with little attention paid to the underlying statistical model. Exponential smoothing equations were merely plausible means at estimating time series features and extrapolating them. There was no way to estimate confidence limits properly, since they depend upon the underlying statistical model.

Some software developers responded to the need for confidence limits with little or no theoretical justification. While the point estimates from such software have been good, the confidence limits have been nearly unusable.

Forecast Pro takes a more modern approach to exponential smoothing. Each variant of exponential smoothing is based upon a formal statistical model which also serves as a basis for computation of confidence limits. The actual smoothing equations are based upon the Kalman filter for the formal statistical model. Of course, all of this is *under the hood*, and you need not know the details.

Models of the Exponential Smoothing Family

Here we will provide an overview—without equations—of the models that make up the exponential smoothing family.

Every exponential smoothing model involves at least the following three components.

Level

Random events

Noise

Simple exponential smoothing involves *only* these components. The data are assumed to consist of the level, slowly and erratically changing as random events impact it, and corrupted by noise. Simple exponential smoothing cannot capture the effects of seasonality or trending.

The remaining components

Trend

Seasonal indexes

Event indexes

are optional. They model features that may or may not be present in the data.

The *trend* can enter in four ways—none, linear, damped or exponential.

The forecasts from an *untrended* model are flat, except perhaps for the effects of seasonal or event indexes.

The forecasts from a *linear* trend model extrapolate the last estimate of the trend without limit. The forecasts eventually become positively or negatively infinite.

The forecasts from a *damped trend* begin almost linearly but die off exponentially until they reach a constant level This may be appropriate for data influenced by business cycles. Damped trend models produce forecasts that remain finite.

The forecasts from an *exponential trend* begin almost linearly but increase as a percentage of themselves. This explosive growth model should only be used when the data are truly growing exponentially.

The Holt model includes a linear trend but does not accommodate seasonal or event effects. The level of the data changes systematically because of the trend. It is also impacted by random events. The trend varies randomly from point to point as it too is impacted by random events. Observations are obscured by noise.

Seasonal indexes can enter in three ways—none, additive or multiplicative.

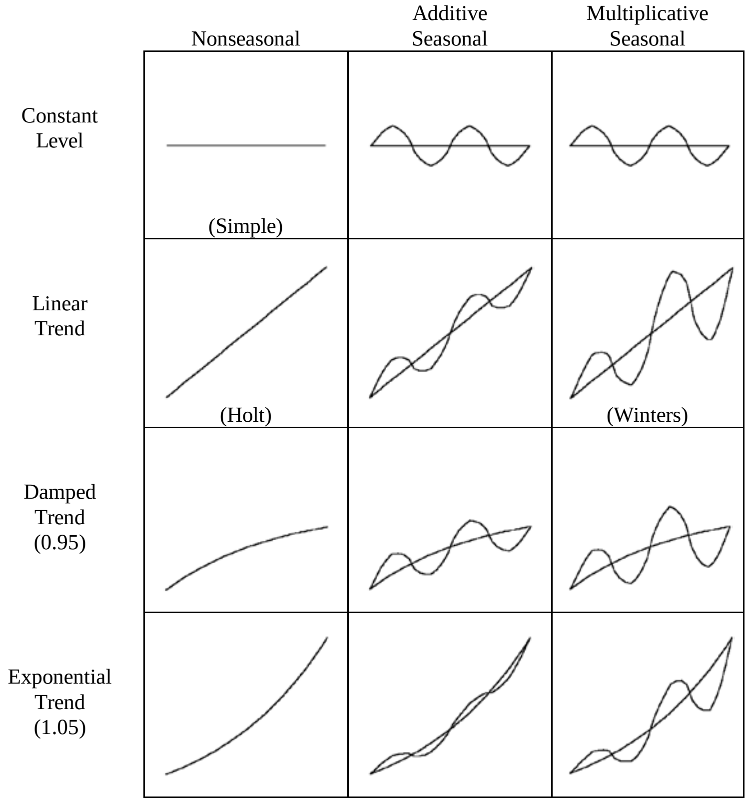
If the indexes are multiplicative, the seasonal adjustment is made by multiplying the index into the deseasonalized series. Thus the effect is proportional to the level of the time series. December sales are adjusted upwards by 20% if the seasonal index is 1.2. This is the most common form of seasonality but it applies only to positive, ratio scale data.

If the indexes are additive, the seasonal adjustment is made by adding the index onto the deseasonalized series. Thus the effect is independent of the level of the time series. December sales are adjusted upwards by 1000 if the seasonal index is 1000.

The multiplicative (additive) Winters exponential smoothing model extracts the level, trend, and multiplicative (additive) seasonal indexes. The underlying nonseasonal model is the same as Holt.

Event indexes can also enter in three different ways—none, additive or multiplicative. The adjustments are analogous to those for seasonal indexes. The difference is that the adjustment is made each time a certain event occurs rather than tying the adjustment to the calendar.

Event index models extend the Holt-Winters family of exponential smoothing models, which includes only the four trend options and three seasonality options, or twelve models in all. The following figure portrays the forecast profiles of these twelve models.



Forecast Profiles of Exponential Smoothing Models (Gardner [1985])

These forecast profiles are created by extrapolating the level, trend and seasonality index estimates from the end of the historic data. They depict the underlying patterns of the data as these patterns exist at the end of the data. They do not and cannot include the effects of future random events or noise, so they are much smoother than the actual future will turn out to be.

Exponential smoothing works as its name suggests. It extracts the level, trend and seasonal indexes by constructing smoothed estimates of these features, weighting recent data more heavily. It adapts to changing structure, but minimizes the effects of outliers and noise.

The degree of smoothing depends upon parameters that must be fitted to the data. The level, trend, seasonal index and event index estimations require one parameter each. If the trend is damped (or exponential), the damping (or growth) constant must also be estimated. The total number of parameters that must be fitted to the data depends on the components of the model.

Implementation of Exponential Smoothing in Forecast Pro

This section presents some details about the Forecast Pro implementation of exponential smoothing.

Model selection

To select a smoothing model automatically, Forecast Pro tries all of the “standard” Holt-Winters candidate models and chooses the one that minimizes the Bayesian information criterion (BIC). The BIC is a goodness-of-fit criterion that penalizes more complex models, i.e., those that require fitting more parameters to the data. Research has shown that this leads to the model that is likely to forecast most accurately (Koehler and Murphree [1986]).

To determine the standard candidate models, Forecast Pro applies the following rules:

1. Automatic model selection does not consider exponential trend models due to their ability to grow explosively in the forecast period. If you wish to build exponential trend models you must use the custom modeling option.

2. If there are less than 5 data points, then Forecast Pro does not attempt to fit a Holt-Winters model to the data. A simple moving average model, which does not require parameter estimation, is substituted.

3. If there is less than two years worth of data, then Forecast Pro Unlimited does not consider seasonal models.

4. If the data contain negatives or zeroes, multiplicative index models are not considered.

If the NA-CL model is under consideration (by default it is) and/or if seasonal simplification models are under consideration (by default they are not). Then an out-of-sample test is used to select amongst the standard model that minimized the BIC and the NA-CL and/or seasonally simplified models.

Parameter optimization

To estimate model parameters, the program uses an iterative search (simplex method) to minimize the sum of squared errors over the historic data. The search begins at default values set by the program. Theoretically, the search could yield a local, rather than the global, minimum. In practice, the authors know of almost no instances where this has occurred or where the algorithm has failed to converge.

Confidence limits

Forecast Pro outputs lower and upper confidence limits for exponential smoothing forecasts. The confidence limits for nonseasonal and additive seasonal models are computed by making the assumption that the underlying probability model is the specific Box-Jenkins model for which the exponential smoothing model is known to be optimal (see Yar and Chatfield [1990]).

The confidence limits for multiplicative seasonal models are computed as described by Chatfield and Yar [1991]. The error standard deviation is assumed to be proportional either (1) to the corresponding seasonal index or (2) to the corresponding seasonal index and the current estimate of the level.

For the nonseasonal models, the error standard deviation is assumed either (1) constant or (2) proportional to the current estimate of the level. For the additive seasonal models, it is assumed either (1) constant or (2) proportional to the current estimate of the seasonalized level.

In each case, Forecast Pro decides which option to use by determining which fits the historical data more closely.

These confidence limits are useful guides to expected model performance, but they are not perfect, since the actual underlying probability model of the data is not known. Their usefulness for multiple-step forecasts deteriorates when the historical errors appear to be correlated.

Notice that the Chatfield-Yar confidence limits differ somewhat from those based on the underlying Box-Jenkins models.

Statistical Description of Exponential Smoothing

Each of the smoothing techniques uses recursive equations to obtain smoothed values for model components. Simple uses one equation (level), Holt uses two (level and trend), Winters uses three (level, trend and seasonal). Event index models require an additional equation. Each equation is controlled by a smoothing parameter. When this parameter is large (close to one), the equation heavily weights the previous values in the series—i.e., the smoothing process is highly adaptive. If the parameter is small (close to zero), the equation weights previous values decreasingly far into the past—i.e., the smoothing process is not highly adaptive.

The following table defines the notation that will be used in the detailed discussion of exponential smoothing. It is adapted from that of Gardner [1985].

m	Forecast lead time
p	Number of periods per year
Y_t	Observed value at time t
S_t	Smoothed level at end of time t
T_t	Smoothed trend at end of time t
I_t	Smoothed seasonal index at end of time t
J_t	Smoothed event index at end of time t
α	Smoothing parameter for level of series
γ	Smoothing parameter for trend

δ	Smoothing parameter for seasonal indexes
λ	Smoothing parameter for event indexes
φ	Damped/exponential trend constant
$\hat{Y}_t(m)$	Forecast for time *t+m* from base t
$\tilde{I}_{t+m}$	Most recent seasonal index for time *t+m*
$\tilde{J}_{t+m}$	Most recent event index for time *t+m*

The Forecast Pro output calls α the *level parameter*, γ the *trend parameter*, δ the *seasonal parameter*, λ the *event parameter* and φ the *decay/growth constant*.

General Additive Index Model

There are twelve exponential smoothing models, so it would not be practical or interesting to discuss each individually. We will instead discuss the most fully featured model and how it relates to simpler models.

The most complex additive index model involves the level S_t, the trend T_t, the seasonal index I_t and the event index J_t. The trend is assumed to decay at the rate $\varphi \leq 1$. The observations Y_t are assumed to be composed of these components as follows.

$$\mathrm{Y_t = S_t + I_t + J_t + e_t}$$

The components S_t, I_t and J_t in this equation are the *true values* for the level, seasonal and event indexes at the time *t*. However, they cannot be observed directly but, rather, must be estimated from the data. This done by using the following recursive equations, which comprise an approximate Kalman filter for the underlying model. The italicized symbols now refer to *estimates* of the true values.

$$S_t = \alpha(Y_t - \tilde{I}_t - \tilde{J}_t) + (1-\alpha)(S_{t-1} + \varphi T_{t-1})$$

$$T_t = \gamma(S_t - S_{t-1}) + (1-\gamma)\varphi T_{t-1}$$

$$I_t = \delta(Y_t - S_t - \tilde{J}_t) + (1-\delta)\tilde{I}_t$$

$$J_t = \lambda(Y_t - S_t - \tilde{I}_t) + (1-\lambda)\tilde{J}_t$$

The symbol $\tilde{I}_t$ refers to the most up-to-date prior estimate of the seasonal index for the month (quarter, week) that occurs at time t. If t refers to December, 1993, then this estimate will have been last updated in December, 1992. The symbol $\tilde{J}_t$ refers to the most up-to-date prior estimate for an event of the type that occurs at time t. These equations update the prior estimates S_{t-1}, T_{t-1}, $\tilde{I}_t$ and $\tilde{J}_t$ to reflect the last observation. The posterior estimates are the quantities on the left hand side of the equations—S_t, T_t, I_t and J_t.

All the simpler additive models are, in a sense, contained in these equations.

> If there is no event at time t, or if event indexes are not wanted, then $J_t = \tilde{J}_t = 0$ and the last equation is discarded.
>
> These equations involve a decaying trend. In this case the decay constant φ is usually a little less than one. To convert the model to a linear trend model, just set φ to 1.0. This is equivalent to erasing it from the equations. To convert the model to an exponential trend model, just set φ to a value greater than 1.0.
>
> If seasonal indexes are not wanted, discard the third equation and set S_t to 0 elsewhere.
>
> If a trend is not wanted, discard the second equation and set T_t to 0 elsewhere.

These equations clearly show how exponential smoothing actually works. Let us look carefully at the first. The quantity $Y_t - \tilde{I}_t - \tilde{J}_t$ represents the current observation, adjusted for seasonal and event effects by subtracting off their last available prior estimates. The adjustment yields an estimate of the current level. The quantity $S_{t-1} + \varphi T_{t-1}$ represents the forecast of the current level S_t based on information available previous to the last observation. The first term, based on the current observation, is weighted by α and the second, based on previous information, is weighted by (1-α).

Each smoothed estimate of the level is computed as a weighted average of the current observation and past data. The weights decrease in an exponential pattern. The rate of decrease depends on the size of the smoothing weight α, which thus controls relative sensitivities to newer and older historic data. The larger the value of the smoothing parameter, the more emphasis on recent observations and the less on distant.

The parameters γ, φ, δ and λ are fitted to the data by finding the values that minimize the sum of squared forecast errors for the historic data. To compute the sum of squared errors for trial values of γ, φ, δ and λ, the following steps are performed.

> The initial values of the four components S_0, T_0 ,I_0 and J_0 are set equal to reasonable guesses based on the data.
>
> The one-step forecast for the first data point t=1 is generated via the equation $\hat{Y}_0(1) = S_0 + \varphi T_0 + \tilde{I}_1 + \tilde{J}_1$. The forecast error $Y_1 - \hat{Y}_1(1)$ is computed and squared.
>
> This step is repeated for t=2 to the end of the historic data t=T. The forecast formula is $\hat{Y}_t(1) = S_t + \varphi T_t + \tilde{I}_{t+1} + \tilde{J}_{t+1}$ so the error is $Y_t - S_t - \varphi T_t - \tilde{I}_{t+1} - \tilde{J}_{t+1}$. As each point is forecasted, the forecast error is squared and accumulated.

This procedure is iterated with new trial values of the parameters until the values that minimize the sum of squared errors are found. The trial parameter values are determined by the simplex procedure, an especially stable algorithm for nonlinear minimization.

Once the parameters have been estimated by fitting to the data, the model is used to compute the forecasts. The equation for the forecast of Y_{T+m} from the forecast base Y_T (last historic data point) is as follows.

$$\hat{Y}_T(m) = S_T + \left[\sum_{i=1}^{m} \varphi^i\right] T_T + \tilde{I}_{T+m} + \tilde{J}_{T+m}\,.$$

General Multiplicative Index Model

The general multiplicative model looks almost the same as the additive, except that multiplication and division replace addition and subtraction. The multiplicative equations are as follows.

$$S_t = \alpha \frac{Y_t}{\tilde{I}_t\ \tilde{J}_t} + (1-\alpha)(S_{t-1} + \varphi\, T_{t-1})$$

$$T_t = \gamma\,(S_t - S_{t-1}) + (1-\gamma)\,\varphi\, T_{t-1}$$

$$I_t = \delta \frac{Y_t}{S_t\ \tilde{J}_t} + (1-\delta)\,\tilde{I}_t$$

$$J_t = \lambda \frac{Y_t}{S_t \tilde{I}_t} + (1-\lambda)\,\tilde{J}_t$$

Simpler models are obtained from these equations in much the same way that they are for the additive case.

> If there is no event at time t, or if event indexes are not wanted, then $J_t = \tilde{J}_t = 1.0$ and the last equation is discarded.
>
> These equations involve a decaying trend. In this case the decay constant φ is usually a little less than one. To convert the model to a linear trend model, set φ equal to 1.0 or simply remove all references to φ.
>
> If seasonal indexes are not wanted, discard the third equation and set to 1.0 elsewhere.
>
> If a trend is not wanted, discard the second equation and set T_t to 0 elsewhere.

Now that the full additive and multiplicative smoothing equations have been presented, we will examine some of the simpler models that they contain as special cases.

Simple Exponential Smoothing

The simple exponential smoothing model is used for data that are untrended, nonseasonal and not driven by promotional events. We can get its equation from either the general additive or general multiplicative model by discarding the last three equations and eliminating the seasonal and event indexes from the first. We are left with the following.

$$S_t = \alpha Y_t + (1-\alpha) S_{t-1} \tag{1}$$

Notice that when

$$\alpha = 1.0$$

the equation becomes

$$S_t = Y_t$$

i.e., there is no “memory” whatsoever of previous values. The forecasts from this model would simply be the last historic point. On the other hand, if the parameter is very small, then a large number of data points receive nearly equal weights, i.e., the memory is long. The other exponential smoothing models use additional smoothing parameters in equations for smoothed values of trend and seasonality, as well as level. These have the same interpretation. The larger the parameter, the more adaptive the model to that particular time series component.

Equation (1) shows how the smoothed level of the series is updated when a new observation becomes available. The m step forecast using observations up to and including the time t is given by

$$\hat{Y}(m) = S_t \tag{2}$$

i.e., the current smoothed level is extended as the forecast into the indefinite future. Clearly, simple exponential smoothing is not appropriate for data that exhibit extended trends.

Holt Exponential Smoothing

Holt’s [1957] exponential smoothing model uses a smoothed estimate of the trend as well as the level to produce forecasts. The forecasting equation is

$$\hat{Y}(m) = S_t + mT_t \tag{3}$$

The current smoothed level is added to the linearly extended current smoothed trend as the forecast into the indefinite future.

The smoothing equations are

$$S_t = \alpha Y_t + (1-\alpha)(S_{t-1} + T_{t-1}) \tag{4}$$

$$T_t = \gamma(S_t - S_{t-1}) + (1-\gamma)T_{t-1} \tag{5}$$

where the symbols were defined previously. Equation (4) shows how the updated value of the smoothed level is computed as the weighted average of new data (first term) and the best estimate of the new level based on old data (second term). In much the same way, equation (5) combines old and new estimates of the one period change of the smoothed level, thus defining the current linear (local) trend.

Multiplicative Winters

In multiplicative Winters, it is assumed that each observation is the product of a deseasonalized value and a seasonal index for that particular month or quarter. The deseasonalized values are assumed to be described by the Holt model. The Winters model involves three smoothing parameters to be used in the level, trend and seasonal index smoothing equations.

The forecasting equation for the multiplicative Winters model is

$$\hat{Y}(m) = (S_t + mT_t)\hat{I}_t(m) \tag{6}$$

i.e., the forecast is computed similarly to the Holt model, then multiplied by the seasonal index of the current period.

The smoothing equations are obtained from the general multiplicative equations by setting φ to 1 and discarding the parts that involve event indexes.

$$S_t = \alpha \frac{Y_t}{I_{t-p}} + (1-\alpha)(S_{t-1} + T_{t-1}) \tag{7}$$

$$T_t = \gamma(S_t - S_{t-1}) + (1-\gamma)T_{t-1} \tag{8}$$

$$I_t = \delta \frac{Y_t}{S_t} + (1-\delta)I_{t-p} \tag{9}$$

The level smoothing equation (7) is similar to equation (4) for the Holt model, except that the latest measurement is deseasonalized by dividing by the seasonal index calculated one year before. The trend smoothing equations of the two models are identical. The seasonal index is estimated as the ratio of the current observation to the current smoothed level, averaged with the previous value for that particular period.

Additive Winters

In additive Winters, it is assumed that each observation is the sum of a deseasonalized value and a seasonal index. The deseasonalized values are assumed to be described by the Holt model. The equations for additive Winters are nearly identical to those of multiplicative, except that deseasonalization requires subtraction instead of division.

The forecasting equation for the additive Winters model is

$$\hat{Y}_t(m) = S_t + mT_t + \hat{I}_t(m) \tag{10}$$

The smoothing equations are obtained from the general additive equations by setting φ to 1 and discarding the event indexes.

$$S_t = \alpha(Y_t - I_{t-p}) + (1-\alpha)(S_{t-1} + T_{t-1}) \tag{11}$$

$$T_t = \gamma(S_t - S_{t-1}) + (1-\gamma)T_{t-1} \tag{12}$$

$$I_t = \delta(Y_t - S_t) + (1-\delta)I_{t-p} \tag{13}$$

NA-Constant Level Model

Forecast Pro supports a form of exponential smoothing that we refer to as an NA-Constant Level model or a “salt” model (a word play on NaCl). The model is a variation of the Additive Winters whereby the trend term is omitted and the smoothing weight α is set to a very small value. By constraining α to a small value the model enforces a constant level and the seasonal component models the departures from the constant level. The model works particularly well for data that exhibit a “selling season” whereby the majority of the demand occurs at specific times of the year (e.g., snow shovels, flu vaccines, etc.).

Seasonally Simplified Exponential Smoothing

In a standard seasonal exponential smoothing model the number of seasonal indices used to model the seasonality equals the number of periods per year. For example there would be 12 seasonal indices for monthly data, 52 seasonal indices for weekly data, etc. Forecast Pro supports seasonally simplified models whereby the number of indices used to model the seasonal component is less than the number of periods per year.

The model is particularly useful for modeling noisy weekly data sets, where using 52 seasonal indices can sometimes result in an overly complex seasonal component that fits to the noise rather than capturing the underlying seasonal pattern.

A seasonally simplified exponential smoothing model substitutes a carefully constructed event model for the seasonal component. The event schedule is constructed to map each period of the year into a seasonal "bucket." For example if you wanted to reduce the number of seasonal indices for a weekly data set from 52 to 26 you would map weeks 1 and 2 of each year into event type 1, weeks 3 and 4 into event type 2, etc. The resulting model would calculate 26 event indices to capture the seasonal pattern. We would refer to the model as having a "bucket size" of 2 since two weeks each year were put into each of the 26 buckets. If we used a bucket size of 4 then the weeks 1-4 of each year would map into event type 1, weeks 5-8 into event type 2, etc. and the resulting model would calculate 13 event indices to capture the seasonal pattern.

Forecast Pro includes the ability to consider seasonally simplified models as part of expert selection as well as allowing you to build them as a customized exponential smoothing model.

Discrete Distributions

Most statistical forecasting models are based on interval data, i.e., data for which zero has no special meaning. Forecasts and data can be negative as well as positive, and the interval from zero to one is statistically equivalent to the interval from 100 to 101. Very little business data are interval in nature but, for the most part , interval data forecast models still perform well.

But there are exceptions. For instance the data might consist entirely of zeroes and small integers. Infrequently used spare parts often fall into this class. The forecasts from simple exponential smoothing for such items may be perfectly reasonable and useful, but the confidence limits are usually unusable.

This is due to the confidence limits from a standard model being symmetric. They do not take into account that sales of these types of items cannot go negative but might become very large. The discrete distributions forecast model produces the same point forecasts but produces much more accurate confidence limits.

Forecast Pro tries two different discrete distributions to fit the data—the Poisson distribution and the negative binomial distribution. Forecast Pro selects the distribution that fits the data better and uses that distribution to compute the forecasts.

Poisson Distribution

The Poisson distribution ranges over integers in the range {0,1,2,...}. It applies to such processes as the number of customers per minute who arrive in a queue, the number of auto accidents per month on a given road, or sales of a particular spare part per month.

The probability that exactly x events occur is given by the following formula.

$$f(x) = \frac{e^{-\lambda}\lambda^{x}}{x!}$$

The Poisson distribution has a single parameter λ that equals both the mean number of events per unit of time, and the variance around the mean. This parameter λ is a positive real number. Forecast Pro chooses the Poisson distribution when the ratio of the sample mean to the sample variance is near unity.

It is likely that the mean number of events per unit of time is actually changing over time. Therefore we must estimate λ as a time series in its own right. It has been shown by Harvey [1989] that this is optimally done via simple exponential smoothing. The current estimate of the level is also an estimate of the current value of λ.

Therefore Forecast Pro performs the following steps.

Use simple exponential smoothing to estimate and forecast λ. The forecasts are equal to the value of λ at the end of the series.

Use the final value of λ to determine from the equation for the distribution the probabilities of 0, 1, 2, ... events per unit of time. These in turn are used to compute integer confidence limits.

The advantage to using a discrete distribution is *not* an improved point forecast but improved confidence limits, and the availability of a formula to compute the probability of zero events, one event, etc.

Negative Binomial Distribution

The variance of many integer series runs higher than can be modeled by the Poisson distribution, where the ratio of the variance to the mean is unity. For instance, the mechanical failures that require a certain spare part may be accurately modeled by a Poisson distribution, but orders for the part may not reflect current failures. The parts may be inventoried by an intermediate distributor, or the end user may buy more than is needed immediately. The result is that, while the mean of the parts orders matches the mean of the failures, the orders are more variable. The negative binomial distribution allows us to model such data.

The negative binomial distribution ranges over the integers {0,1,2,...}. The probability that exactly x events occur is given by the following formula.

$$f(x) = \frac{(x+y-1)!}{x!(y-1)!} p^y (1-p)^x$$

This is the formula for the probability of x failures before the y^{th} success in a sequence of Bernoulli trials in which the probability of success at each trial is p. In Forecast Pro, we regard the negative binomial distribution in a more empirical way. It is flexible enough to model many discrete business series that are not modeled well by the Poisson distribution.

The parameters to be fitted to the data consist of y, an integer which assumes values in the range {1,2,3,...}, and p, which lies in the interval from 0.0 to 1.0. These two parameters are fitted to the data by using the facts that the mean of the distribution is $y(1-p)/p$ while its variance is $y(1-p)/p^2$. Thus the ratio of the mean to the variance is p. The mean of the series is estimated via simple exponential smoothing, as with the Poisson distribution. We assume that the ratio of the variance to the mean is a constant, which we also estimate from the data.

This gives us the distribution of x at the end of the historic data. The point forecasts equal the final estimate of the mean. The confidence limits are computed by using the formula for f(x) as a function of x and the two fitted parameters.

Croston's Intermittent Demand Model

Description

Much product data, especially for lower volume items, are intermittent in nature. For many, or even most periods, there is no demand at all. For periods where there is some demand, it is randomly distributed independently or nearly independently of the demand interval. This might be the case for spare parts that are usually ordered in batches to replenish downstream inventories. The Poisson and negative binomial distributions do not usually fit such data well because they link the zeroes and non-zeroes as part of the same distribution.

Croston [1972] proposed that such data be modeled as a two-step process. He assumed that the *demand intervals* were identically independently distributed (iid) geometric. This is equivalent to assuming that the probability that a non-zero demand occurs in any given period is iid Bernoulli, as though by the flip of an unfair coin. He further assumed that the distribution of successive demands was iid normal.

The alternative model for data this messy is usually simple exponential smoothing. This yields horizontal forecasts at a level fitted adaptively to the data. Willemain et al. [1994] examined the performance of a variant of the Croston model relative to that of exponential smoothing, and found it markedly superior in forecast accuracy, both for simulated and real data, even in the presence of autocorrelation and cross-correlation between demand size and demand interval.

The variation that Willemain et al. introduced was the substitution of the log normal distribution for the normal distribution for successive order sizes. This is sensible for most data because the probability of non-positive demand size is zero. However, it cannot be applied to demand data that sometimes goes negative, as it sometimes does when a company registers returns as negative demand.

Implementation

Two basic models are implemented in Forecast Pro—the Croston Model as originally implemented and the Willemain variant. The Willemain variant is always selected unless there are occasional negatives in the historic data. If numerous data points are negative, the Croston model is tagged as inappropriate, as it also is when the vast majority of the historic periods have a non-zero demand.

The quantities that must be estimated from the data include the following.

Probability of a demand in a given period, adaptively estimated to reflect conditions at the end of the historic data via simple exponential smoothing of the demand interval. The smoothing parameter is optimized, as is that for the mean order size, by minimizing the sum of squared fitting errors.

Mean order size, adaptively estimated in the same way.

Standard deviation of the demand size, estimated globally over the historic data set.

The forecasts are computed as the product of demand probability and demand size. All three of the estimated quantities are used to compute the overall distribution function, from which the confidence intervals are computed.

Curve Fitting

Curve fitting is generally used to model the global trend of a data set. Although curve fitting is not as sophisticated as some of the other Forecast Pro forecasting methodologies, it can still be quite useful. Unlike some of the other methods, curve fitting may be used with short time series (the suggested minimum length is ten data points). Also, the program provides a quick and easy way to identify the general form of the curve your data are following. Be aware however, that curve fitting methods do not accommodate for or project seasonal patterns in a series.

The curve fitting routine supports four types of curves—straight line, quadratic, exponential and growth (s-curve). You can let Forecast Pro choose the form of the curve or select it yourself.

The automatic tries the four curves and selects the one that minimizes the BIC for the historic series. The equations for each curve are shown below (t=time). All of the coefficients of the model are selected to minimize the sum of the squared errors.

Straight line: $\mathrm{Y} = a + bT$

Quadratic: $\mathrm{Y} = a + bT + cT^2$

Exponential: $\mathrm{Y} = e^{a+bT}$

Growth Curve: $\mathrm{Y} = \dfrac{a}{1 + e^{-b(T-c)}}$

Box-Jenkins Statistical Models

Box-Jenkins is a powerful forecasting technique which, for appropriate data, often outperforms exponential smoothing. Traditionally, however, Box-Jenkins models have been difficult and time consuming to build. This has kept them from widespread acceptance in the business community.

However, automatic algorithms such as those found in Forecast Pro, now allow forecasters to build Box-Jenkins models quickly and easily. As a result, they are now candidates for more widespread use.

In the largest empirical studies of forecasting accuracy to date (Makridakis [1982], Makridakis [2000]), exponential smoothing outperformed Box-Jenkins *overall*, but in many specific cases, Box-Jenkins outperformed exponential smoothing. Ideally, a forecaster would switch between Box-Jenkins and exponential smoothing models, depending on the properties of the data. This is precisely what the Forecast Pro expert system is designed to do.

Box-Jenkins models are built directly on the autocorrelation function (ACF) of a time series variable. Therefore, a prerequisite for Box-Jenkins is that the data should possess a reasonably stable autocorrelation function. If the autocorrelations are not stable, or the data are too short (say, fewer than 40 points) to permit reasonably accurate estimates of the autocorrelations, then exponential smoothing is the better choice. This avoids the principal pitfall of Box-Jenkins: fitting a complex model to chance historic correlations or outliers.

Univariate Box-Jenkins cannot exploit leading indicators or explanatory variables. If these are important, then a multivariate method such as dynamic regression is a better choice.

Forecast Pro implements the univariate ARIMA (AutoRegressive Integrated Moving Average) procedure described by Box and Jenkins [1976]. The models can be identified completely automatically by the program, or the user can interactively build a model, or test variations on the model selected by the Forecast Pro expert system. The program supports the multiplicative seasonal model described by Box and Jenkins.

This section is intended to provide a background to the statistical methodology used in the program. Those who would like a more thorough coverage should consult Box and Jenkins' classic theoretical textbook or *Applied Statistical Forecasting* [Goodrich 1989].

Implementation of Box-Jenkins in Forecast Pro

Automatic identification

The program begins by performing a range-mean test to determine whether the log or square root transform should be applied to the data.

Next the program determines the simple and seasonal differencing necessary to render the data stationary. It uses an adaptation of the Augmented Dickey-Fuller test (see Goodrich [1989]). Then it computes approximate values for the parameters of a group of candidate models. Forecast Pro tests each model, and selects the one that minimizes the BIC criterion.

Exhaustive fitting and examination of all low order ARIMA models would take an inordinate amount of computer time. Forecast Pro actually overfits a *state space* model, and uses it to generate approximate Box-Jenkins models quickly. Sometimes this method misses the minimum BIC by a slight amount, but it virtually never selects a bad model.

Business Forecast Systems has compared its Automatic Box-Jenkins models with the published results from the M-competition, where an expert spent 20 minutes to identify each ARIMA model manually. Forecast Pro outperformed the Box-Jenkins expert at every forecast horizon.

Business Forecast Systems recommends that you use Automatic identification routinely. Use Custom identification only when the program so suggests, or when you have a strong reason to reject the automatic model.

Initialization

Forecast Pro uses the method of back-forecasting to initialize Box-Jenkins models. This technique is described in Box and Jenkins [1976].

Parameter estimation

Forecast Pro uses the method of unconditional least squares to obtain final parameter estimates. If necessary, the parameters are adjusted to ensure stationarity or invertibility

Constant term

By default, Forecast Pro uses a constant term only when an ARIMA model does not involve differencing. This is to avoid imposition of deterministic trends, which can lead to large forecast errors at longer horizons. You can, however, override the default if you want. In that case Forecast Pro will estimate the constant as though it were another parameter, so that you can check its statistical significance.

Box-Jenkins Background

Two statistical concepts are pivotal for understanding the Box-Jenkins modeling and dynamic regression, stationarity and autocorrelation.

Stationarity

A time series process is stationary (wide sense) when it remains in statistical equilibrium with unchanging mean, unchanging variance and unchanging autocorrelations. A stationary process can be represented as an optimal autoregressive moving average (ARMA) model.

Unfortunately, most business and economic time series are not stationary. There are many forms of nonstationarity, but the following forms are especially important.

Nonstationarity in the mean.

The mean is not constant but drifts slowly, without consistent direction.

The time series is trended or cyclical. The trend is not constant but slowly drifts up and down.

Nonstationarity in the variance.

The time series is heteroscedastic, i.e. the variance of observations around the mean is changing.

One treats these cases by transforming the data to stationarity. Nonstationarity in the mean is removed by *differencing*. Nonstationarity in the variance is removed by applying a *Box-Cox power transform*.

Autocorrelation function

According to ARIMA statistical theory, a time series can be described by the joint normal probability distribution of its observations Y_1, Y_2, ... , Y_N. This distribution is characterized by the vector of *means* and the *autocovariance* function.

The autocovariance of Y_t and its value Y_{t+m} at a time m periods later is defined by

$$\gamma_m = \mathrm{cov}(Y_t, Y_{t+m}) = E\left[(Y_t - \mu)(Y_{t+m} - \mu)\right],$$

where the operator E denotes statistical expectation, cov denotes the covariance, and μ is the expected value of Y_t. Notice that the autocovariance function is a function of the time *separation* m, not the absolute times. This is an implicit assumption that the autocovariance function does not depend on the time origin t. In other words, the time series is *stationary*. If it is not, then its autocovariance function is not defined.

Notice that γ_0 is the same as the variance σ_Y^2. The autocorrelation function is computed by dividing each term of the autocovariance function by the variance σ_Y^2:

$$\rho_m = \frac{E(Y_t - \mu)(Y_{t+m} - \mu)}{\sigma_Y^2}$$

The autocovariance function is a *theoretical* construct describing a statistical distribution. In practice, we can only obtain *estimates* of the true values. The generally accepted formula is

$$\rho_m = E\left[\frac{1}{T}\sum_{t=1}^{T-m}\left(Y_t - \overline{Y}\right)\left(Y_{t+m} - \overline{Y}\right)\right]$$

where $\overline{Y}$ is the *sample* mean. The sample autocorrelation function is then given by

$$r_m = \frac{c_m}{c_o}$$

The sampling error of this estimate can be large, especially when the autocorrelations are themselves substantial. The estimates are also highly intercorrelated. Because of this, one must use caution in labeling particular correlations significant by visual examination of the sample autocorrelation function.

The sample autocorrelation function displayed in Forecast Pro includes dashed lines at the 2σ limits, where σ is the approximate standard error of the sample autocorrelation coefficient, computed via the Bartlett [1946] approximation. The rate at which σ expands depends on the sample values of lower order autocorrelations.

Description of the ARIMA Model

Box-Jenkins[1] models the autocorrelation function of a stationary time series with the minimum possible number of parameters. Since the Box-Jenkins dynamic model includes features (moving average terms) that dynamic regression does not, Box-Jenkins theoretically will produce the optimum univariate dynamic model. Therefore, even when a dynamic regression model might ultimately be selected, a preliminary Box-Jenkins analysis provides a useful benchmark for model dynamics. Since the procedure is quick and automatic, this puts very little analytic burden on the user.

The Box-Jenkins model uses a combination of autoregressive (AR), integration (I) and moving average (MA) terms in the general AutoRegressive Integrated Moving Average (ARIMA) model. This family of models can represent the correlation structure of a univariate time series with the minimum number of parameters to be fitted. Thus these models are very efficient statistically and can produce high performance forecasts.

The notation we will use is consistent with that used for exponential smoothing.

N	Number of historic data points
m	Forecast lead time (horizon)
p	Number of periods in a year
Y_t	Observed value at time t

[1]Properly, Box-Jenkins refers to the *modeling procedure* that these two statisticians devised to fit *ARIMA* models to data, and not the model itself. In this document, however, we use the two terms almost interchangeably.

∇_t	Differencing operator
∇_s	Seasonal differencing operator
B	Backward shift operator
ϕ_i	Autoregressive coefficient (lag i). In Forecast Pro this term is displayed as *a[i]*.
$\phi(B)$	Autoregressive polynomial of order p
Φ_i	Seasonal autoregressive coefficient (lag i) In Forecast Pro this term is displayed as *A[i]*.
$\Phi(B^S)$	Seasonal autoregressive polynomial of order p_s
θ_i	Moving average coefficient (lag i). In Forecast Pro this term is displayed as *b[i]*.
$\theta(B)$	Moving average polynomial of order q
Θi	Seasonal moving average coefficient (lag i). In Forecast Pro this term is displayed as *B[i]*.
$\Theta(B^S)$	Seasonal moving average polynomial of order q_s
$\hat{Y}_t(m)$	Forecast for time *t+m* from origin *t*
e_t	One-step forecast error $Y_t - Y_{t-1}$
ε_t	Normally independently distributed random shock.

Differencing

If a time series is not stationary in the mean, then the time series must first be transformed to stationarity by the use of differencing transforms. To describe differencing transforms we use the backward shift operator B, defined as follows.

$$BY_t = Y_{t-1}$$

$$B^m Y_t = Y_{t-m}$$

This operator will be used in our discussion of ARMA processes. For instance, the differencing operator is defined as follows.

$$\nabla = (1 - B)$$

Autoregressive processes

The AR(p) model is specified by the equation

$$Y_t - \phi_1 Y_{t-1} - \phi_2 Y_{t-2} - \ldots - \phi_p Y_{t-p} = \varepsilon_t \tag{1}$$

in which the dependent variable appears to be regressed on its own lagged values. This equation can also be represented in terms of the backward shift operator B as

$$(1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p) Y_t = \varepsilon_t \tag{2}$$

and, if we adopt the notation $\phi(B)$ for the polynomial in B, it can be written succinctly in the form

$$\phi(B) Y_t = \varepsilon_t \tag{3}$$

Moving average processes

The Moving Average process MA(q) is given by

$$Y = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \theta_q \varepsilon_{t-q} \tag{4}$$

or, alternatively, in the polynomial form

$$Y_t = \theta(B) \varepsilon_t \tag{5}$$

The pure moving average process MA(q) is virtually never observed in real world data. It describes the unlikely process whose autocorrelations are nonzero for q lags, and zero thereafter.

Moving average terms are, in practice, used only in conjunction with differencing or autoregressive terms. In that case, they are invaluable. They induce data smoothing just like that of exponential smoothing.

ARMA and ARIMA processes

The AutoRegressive Moving Average process ARMA(p,q) combines the features of the AR(p) and MA(q) processes. In polynomial form, it is given by

$$\phi(B) Y_t = \theta(B) \varepsilon_t \tag{6}$$

Thus the AR(p) process is the same as ARMA(p,0) and the MA(q) process is the same as ARMA(0,q).

Any stationary time series can be modeled as an ARMA(p,q) process. Any time series that can be made stationary by differencing d times can be modeled as an ARIMA(p,d,q) process. The ARIMA(p,d,q) model is given by the following equation.

$$\phi(B)(1-B)^d Y_t = \theta(B) \varepsilon_t \tag{7a}$$

This is the most general nonseasonal Box-Jenkins model.

Deterministic trends

By default, Forecast Pro does not include a constant term in an ARIMA model except when the model does not involve differencing. If you dictate that a constant term be used the equation for the model now takes the form shown in equation (7b).

$$\phi(B)(1-B)^d Y_t = \theta(B)\varepsilon_t + c \tag{7b}$$

The effect of a constant term is to introduce deterministic trending into your model, in addition to its other properties. If you have differenced once, the trend is linear; if you have differenced twice, it is quadratic.

This is usually undesirable because it extrapolates the global trend of the historic data indefinitely into the future, even when the current trend is slight. This usually produces poor forecast accuracy for longer horizons. Business Forecast Systems has confirmed this effect by testing over the 111 Makridakis time series.

Seasonal Models

Equation (7a) is adequate to model many seasonal series, provided that the polynomials “reach back” one or more seasonal periods. This means that either p or q (or both) must equal or exceed the seasonal period s. Since all intervening terms would also be included, such a model is not parsimonious, i.e., it would contain unnecessary coefficients to be estimated. This is often damaging to predictive validity of the model.

On the other hand, we might consider a seasonal version of equation (7) in which the backward shift operator *B* is replaced by its seasonal counterpart B^s. The resulting equation is

$$\Phi(B^s)(1-B^s)^D Y_t = \Theta(B^s)\varepsilon_t \tag{8}$$

where the polynomials θ and Θ are of orders P and Q respectively. This is the $ARIMA(P,D,Q)_S$ model. It relates the observation in a given period to those of the same period in previous years, but not to observations in more recent periods.

The most general seasonal model includes both seasonal and simple ARIMA models at once. The following equation describes the *multiplicative seasonal ARIMA* model.

$$\phi(B)\Phi(B^s)(1-B)^d(1-B^s)^D Y_t = \theta(B)\Theta(B^s)\varepsilon_t \tag{9}$$

It is usually symbolized as ARIMA (p,d,q)(P,D,Q).

Selecting Model Orders

The hardest part of Box-Jenkins modeling is deciding *which* ARIMA(p,d,q) model fits the data best, i.e. in identifying the degree of differencing d, the AR order p, and the MA order q. Much of Box and Jenkins [1976] is devoted to this so-called "identification" problem. The Forecast Pro expert system, in fact, identifies the model *automatically* . Therefore, the remainder of this section, in which the *original* Box-Jenkins procedure is presented, is inessential.

The original Box-Jenkins procedure depends upon graphical and numerical analysis of the autocorrelation function and the partial autocorrelation function. It is a pattern recognition procedure that requires skill and patience to learn. We will discuss only the nonseasonal case.

Degree of differencing

The identification procedure begins by determining the degree of differencing d that is required to make the original data Y_t stationary. This is done through examination of the autocorrelation function r_k.

The first few lags of the autocorrelation function of the raw data Y_t are inspected; if these die out relatively quickly, then no differencing is required, i.e. d=0. If not, then the original data are replaced by its first difference ∇Y_t and the process is repeated. If the autocorrelation function of the differenced data dies out quickly, d=1. If not, the data are differenced a second time to obtain $\nabla^2 Y_t$. This process is repeated until, for some d, the autocorrelation function of the multiply differenced data does die out quickly. In practice, d is rarely greater than 2.

Once the degree of differencing is determined, the remainder of the analysis deals with the stationary series $\nabla^d Y_t$. If d is zero, these are the original data.

Autoregressive order

The autoregressive order p is determined by inspection of the sample partial autocorrelation function $\hat{\phi}_{kk}$. We will motivate the odd notation and the definition of this function through a thought experiment.

Suppose that the process is thought to be purely autoregressive (q=0). Then a rational strategy to determine p would be to compute a regression of Y_t on its first lag, then on its first two lags, and so on until the last lag introduced into the regression turns out not to be statistically significant. This is determined by a statistical test on $\hat{\phi}_{kk}$, which is defined as the coefficient of Y_{t-k} in a regression on $Y_{t-1}, Y_{t-2}, \ldots, Y_{t-1}$, i.e. the k'th coefficient of the k'th regression.

Actually, a fast recursive algorithm is used instead of performing so many regressions. A graph is presented of the first forty-eight lags so that the AR order can be determined. If the process is ARIMA(p,d,0), then the partial autocorrelation function dies abruptly after p lags.

Moving average order

A pure moving average process ARIMA(0,d,q) exhibits the same behavior in the *autocorrelation* function that the autoregressive process ARIMA(p,d,0) does in the *partial autocorrelation* function. In other words, if the process is ARIMA(0,d,q), then the sequence r_k is large for k<q+1 and small for k>q. Thus the autocorrelation function is used for MA processes in the same manner as the partial autocorrelation function for AR processes.

The functions r_k and $\hat{\phi}_{kk}$ also exhibit similar behavior for ARIMA(p,d,0) and ARIMA(0,d,q) processes, respectively. Instead of abruptly cutting off at p and q, respectively, these functions tail off smoothly in exponential decay or exponentially damped sine waves. By examining both functions, the forecaster can determine the orders of pure AR and MA processes.

Mixed processes ARIMA(p,d,q) are more complex. Neither the partial autocorrelation function nor the autocorrelation function abruptly dies out. Instead, the autocorrelations remain large for k≤q+1 and die out exponentially thereafter. The partial autocorrelations remain large for k<p+1 and die out for k>p. Manual identification of mixed ARIMA processes is often very difficult.

There are two severe problems with this procedure for order identification.

> First, even when the data really does fit an ARIMA process, the sample autocorrelations used to identify the process can be very different from the theoretical ones due to sampling variation.
>
> Second, the actual data usually contain outliers and other unmodelable features that can significantly distort the autocorrelation and partial autocorrelation functions. It is our judgment that the Box-Jenkins procedure should be used only as the very roughest guide.

We recommend that you fit an automatic Box-Jenkins model first. Then, if you suspect that you can find a better model, you can try variations of the automatic model. You can use the BIC criterion to make a final decision. Note that the Forecast Pro automatic identification method has bested human experts in several academic studies.

Dynamic Regression

Forecast Pro dynamic regression supports the development of forecasts that combine time-series-oriented dynamic modeling and the effects of explanatory variables or leading indicators. The conventional regression model is enhanced by including support for an extension of the Cochrane-Orcutt autoregressive error model, and for the use of lagged dependent and independent variables. Forecast Pro does not support the development of simultaneous equation models.

Dynamic regression should be used when (1) the data are long enough and stable enough to support a correlational model (2) explanatory variables result in a definite increase in accuracy of fit and (3) reliable forecasts for the explanatory variables are available. Remember that complex models often produce forecasts that are less accurate than those from simpler models, even though they may fit the historic data better.

Description of Dynamic Regression Model

The ordinary least squares dynamic regression model takes the form

$$P(B)Y_t = \beta X_t + e_t \tag{1}$$

where the errors e_t are independently identically normally distributed. The symbols in this equation and the equations to follow are defined in the table below.

N	Number of historic data points
M	Forecast lead time (horizon)
s	Number of periods in a year
Y_t	Observed value at time t
X_{it}	Observed value of i’th explanatory variable at time t
B	Backward shift operator
ϕ_i	Autoregressive coefficient of Y_{t-i}
ρ_i	Autoregressive coefficient of e_{t-i}
β_i	Coefficient of X_i
$\hat{Y}_t(m)$	Forecast for time $t+m$ from origin t
e_t	One-step forecast error $Y_t - Y_{t-1}$

The lags of the dependent variable are contained in the polynomial P(B), just as in the Box-Jenkins model. The dynamic regression model differs from Box-Jenkins in two important ways:

> It includes one or more *independent* variables, which *drive* the process. For example, advertising or promotion usually drive sales.

Equation (1) does not include moving average terms, which are often very useful in Box-Jenkins. Regression models will therefore be less parsimonious than Box-Jenkins for some processes.

Thus dynamic regression is stronger than Box-Jenkins in one way and weaker in another.

It will often be found that the errors obtained from equation (1) are correlated, contrary to assumption. This can be determined by examination of diagnostics in the dynamics module. This may indicate that additional lags of the dependent variable should be introduced, or additional independent variables or new lags of existing independent variables should be introduced, or both.

The generalized Cochrane-Orcutt model is an alternative way to improve model dynamics that often requires estimation of fewer new parameters. In the Cochrane-Orcutt model, equation (1) is replaced by the pair of equations

$$P(B)Y_t = \beta X_t + \nu_t \tag{2}$$

$$R(B)\nu_t = e_t \tag{3}$$

in which the raw residuals are correlated via an autoregressive process specified by the polynomial R(B) in the backward shift operator. Equations (2) and (3) can be rewritten as a single equation

$$R(B)P(B)Y_t = R(B)\beta X_t + e_t \tag{4}$$

Dynamic Regression Diagnostics

A regression model is far harder to fit to the historic data than a Box-Jenkins model for several reasons. First, the dynamic portion of the model (lagged dependent variable and Cochrane-Orcutt terms) must be determined term by term on the basis of hypothesis testing, rather than automatically. Second, there are no moving average error terms in dynamic regression; if they are needed they must be approximated by additional complexity in the dynamic regression model. Third, the explanatory portion of the model adds an additional layer of complexity over the univariate case. Moreover, the lag distribution of the explanatory variables must also be considered. As a result there may be hundreds of specific terms that should be considered in a particular model.

This complex situation calls for an orderly and systematic strategy. The Forecast Pro regression diagnostics are modularized into three batteries of tests aimed at two phases of the model development process. These phases are:

Development of the dynamic model

Development of the explanatory model

The dynamic regression test battery provides specific diagnostics for the current model. Most of the diagnostics are chi-squared statistics based on Lagrange multiplier tests

(Engle [1984]). Lagrange multiplier tests are asymptotically equivalent to the more commonly used Wald tests and likelihood ratio tests. The following paragraphs describe the tests.

Each diagnostic tests for a specific deficiency in the model. However, they are not independent of each other. A deficiency in one specific area can cause several other test statistics to become significant as well. Because of this it is best to find the test where the null hypothesis is rejected at the highest probability, and make that one specific change. Then, reexamine the diagnostics for the altered model.

Dynamics specification

The first group of diagnostics tests for inclusion of Cochrane-Orcutt autoregressive error terms. The tests are described below.

> **_AUTO[-n]**. The alternative hypothesis is that an error autocorrelation of lag n should be added to the model. Forecast Pro performs a test for each of the first twelve lags and the first two seasonal lags. A test is omitted if the term is already in the model.

The remaining dynamics tests check for inclusion of lagged dependent variables.

> **Y[-n] test**. The alternative hypothesis is that the n'th lag of the dependent variable should be added to the model. Forecast Pro performs a test for each of the first twelve lags and the first two seasonal lags. It uses the actual name of the variable. A test is omitted if the term is already in the model.

The program recommends that some specific new term be added to the model, unless all tests are insignificant at the level 0.01.

Variable specification

The variable specification tests check for problems in specification of the independent variables. The tests are described below.

> **Excluded variables**. A Lagrange multiplier test is computed for each inactive variable on the script.

> **Time trend**. The alternative hypothesis is that a linear time trend improves the model. A significant test does not necessarily indicate that a time trend variable should be added. The problem often lies with model dynamics or by the exclusion of some other variable.

> **Constant term**. The alternative hypothesis is that a constant term improves the model.

Lagged independent variables. A test is made for each independent variable now present in the model. The alternative hypothesis is that its first lag should also be in the model.

Custom excluded variable tests

The alternative hypothesis in the excluded variables test described above is that the model should include the single additional variable specified. The custom excluded variables test option allows you to test combinations of excluded variables.

It is not uncommon that combinations of variables will be jointly significant even when they are separately insignificant.

Bass Diffusion Model

The Bass diffusion model is a new product forecasting technique that can be used with or without historic demand data. The Bass model is most often used to forecast first time adoptions of new-to-world products.

The model tries to capture the adoption rates of two types of users—innovators and imitators. Innovators are early adopters of new products and are driven by their desire to try new technology. Imitators are more wary of new technology—they tend to adopt only after receiving feedback from others.

$$Y_t = p(m - \sum Y_t) + q(\frac{\sum Y_t}{m})(m - \sum Y_t) \qquad (1)$$

Y_t	Number of adopters at time t
m	Number of potential adopters over entire life cycle
p	Coefficient of Innovation
q	Coefficient of Imitation

The Bass model can be written in several different forms. The form in equation (1) is adapted from Kahn[2006]. Notice the plus sign on the left hand side of the equation separates the innovation component from the imitation component. Conceptually equation (1) can be thought of as:

$$Y_t = (p * \text{Remaining Potential}) + (q * \text{Current Adopters} * \text{Remaining Potential}) \qquad (2)$$

Equation (2) illustrates how *p* defines the strength of the Innovation Effect and *q* defines the strength of the Imitation Effect.

If you have 5 or more historic data points, *p, q* and *m* can be fit to the data using regression. Consult Bass[2004] for details.

With fewer than 5 historic data points , *p*, *q* and *m* must be input into the model. In these instances, the coefficients could be set using values from an analogous product's model. There is also a considerable body of literature on the Bass model including published coefficients for different types of technologies. Consult Lilien, Rangaswamy and Van den Bulte[1999].

Forecasting By Analogy

By Analogy is a new product forecasting technique that can be used with or without historic demand data. The approach is sometimes also referred to as "looks like" analysis.

The concept is a very simple one. You are launching a new product and you expect the initial sales pattern to be similar to an analogous product's initial sales pattern or to a "launch profile" that you've created.

If the product has not yet launched (i.e., there is no historic data available) then you must supply an estimate of the initial sales over a specific period of time (the "launch total" over the "launch horizon"). Forecast Pro will then create the forecast by proportionally allocating the launch total over the launch horizon using the analog series to define the proportions.

If historic data exists, Forecast Pro will calculate and display an "estimated launch total". To do so, it first uses the analog series to determine the cumulative percentage of the launch total that the available historic data represent, it then assumes that the sum of the available history equals that cumulative percentage and estimates the launch total. For example, if there are 5 historic demand observations that sum to 500 and the sum of the first 5 periods of the analog series corresponds to 40% of the analog series' launch total, then 500 is assumed to equal 40% of the estimated launch total and thus the estimated launch total equals 1,250.

If historic data exists and you specify that the estimated launch total should be used to generate the forecast, Forecast Pro will create the fitted values and forecasts by proportionally allocating the estimated launch total over the launch horizon using the analog series to define the proportions.

If historic data exists and you specify a launch total to be used, Forecast Pro will subtract the sum of the available history from the specified launch total to ascertain the cumulative forecast needed so that the sum of the available history and forecast will equal the specified launch total. It then spreads the needed cumulative forecast value using the analog series' forecast values to define the proportions. The same proportionality factors used to generate the forecasts are then used to generate the fitted values—thus the fitted values represent the historic volume that would normally be associated with the forecast.

Model Statistics

Within-sample statistics are displayed each time a model is fitted to the data. Out-of-sample statistics are displayed whenever a hold out sample is used. Each statistic is listed below:

Sample size. The number of historical data points used to fit the model. Operations that discard data points, (e.g., differencing, inclusion of lagged variables, etc.) can reduce this statistic.

Number of parameters. The number of fitted parameters (coefficients) in the model.

Mean. The sample mean (average) for the historical data.

$$\overline{Y} = \frac{1}{n}\sum (Y_t)$$

Standard deviation. A measurement of the dispersion of the historical data around its mean.

$$S = \sqrt{\frac{1}{n-1}\sum (Y_t - \overline{Y})^2}$$

R-square. R-square is the fraction of variance explained by the model.

$$R^2 = 1 - \frac{\sum (Y_t - F_t)^2}{\sum (Y_t - \overline{Y})^2}$$

Adjusted R-square. The adjusted R-square is identical to the R-square except that it is adjusted for the number of parameters (k) in the model.

$$\overline{R}^2 = 1 - (1 - R^2)\frac{n-1}{n-k-1}$$

Durbin-Watson. The Durbin-Watson d-statistic is used to test for correlation of adjacent fitted errors, i.e. for first-lag autocorrelation. If T is the number of sample points and e_t is the fitted error at point t, then d is computed as follows.

$$d = \frac{\sum_{t=2}^{T} (e_t - e_{t-1})^2}{\sum_{t=1}^{T} e_t^2}$$

While the d-statistic is easy to compute, it is hard to interpret.

The null hypothesis is that the first-lag autocorrelation is zero. One looks up the Durbin-Watson bounds d_L and d_U for sample size T and significance α in a table. The null is accepted if $d < d_L$ and rejected if $d > d_U$. If $d_L < d < d_U$, then the test is inconclusive. Our recommendation, with which many disagree, is to reject the null only when the test is conclusive.

Another problem is that the *d*-statistic is not strictly valid for models with lagged dependent variables. In that case, many statisticians use the Durbin *h*-statistic instead. The Durbin *h* is not reported in Forecast Pro.

We recommend that you rely on the Ljung-Box test, which is straightforward, and on visual examination of the error autocorrelation function.

Ljung-Box test. The Ljung-Box Q-statistic, which is used to test for overall autocorrelation of the fitted errors of a model, is a statistical improvement on the Box-Pierce (portmanteau) test. If T is the number of sample points, r_i is the i'th autocorrelation coefficient, and L the number of autocorrelation coefficients, then Q is computed as follows.

$$Q = T(T+2)\sum_{i=1}^{L} \frac{r_i^2}{(T-i)}$$

The statistic is a weighted sum of squared autocorrelations, so it is zero only when every autocorrelation is zero. The more autocorrelation, the greater the size of Q. The weights are selected to make Q approximately $\mathrm{X}^2(L-n)$, i.e. Chi-square with L-n degrees of freedom.

Forecast error. The standard forecast error is the root mean square of the fitted errors adjusted for the number of parameters (k) in the model. It is used to compute the confidence limits of the forecasts, but, realistically, it is usually an overly optimistic estimate of true out-of-sample error.

$$FE = \sqrt{\frac{\sum (Y_t - F_t)^2}{n-k}}$$

BIC. The AIC (Akaike Information Criterion) and the BIC (Bayesian Information Criterion) are the two *order estimation criteria* in most common use. A specific model is selected from a model family by finding the model that minimizes the AIC or BIC.

Either statistic rewards goodness-of-fit, as measured by the root mean square error s, and penalizes for complexity, i.e. the number of parameters n. Koehler and Murphree [1986] showed that, for series from the M-competition, the BIC leads to better out-of-sample forecast performance and, for this reason, Forecast Pro uses and displays the BIC.

There are several equivalent versions of the BIC, related to each other by transforms. In Forecast Pro, we use the following equation, in which T represents the sample size.

$$BIC = sT^{n/2T}$$

This version of the BIC is scaled the same as the standard forecast error. It can very loosely be interpreted as an estimate of out-of-sample forecast error.

The BIC can be used to compare different models from the same family, and for the same data. Since it is scaled to the standard forecast error, it is meaningless as an absolute criterion of merit.

MAPE. The MAPE (Mean Absolute Percentage Error) is used to measure within sample goodness-of-fit and out-of-sample forecast performance. It is calculated as the average of the unsigned percentage errors.

$$MAPE = \frac{1}{n}\sum\frac{|Y_t - F_t|}{|Y_t|}$$

SMAPE. The SMAPE (Symmetric Mean Absolute Percentage Error) is a variation on the MAPE that is calculated using the average of the absolute value of the actual and the absolute value of the forecast in the denominator. This statistic is preferred to the MAPE by some and was used as an accuracy measure in several forecasting competitions.

$$SMAPE = \frac{1}{n}\sum\frac{|Y_t - F_t|}{(|Y_t + F_t|)/2}$$

RMSE. The RMSE (Root Mean Square Error) is used to measure within sample goodness-of-fit. It is calculated as the square root of the average of the squared errors.

$$RMSE = \sqrt{\frac{1}{n}\sum(Y_t - F_t)^2}$$

MAD. The MAD (Mean Absolute Deviation) is used to measure within sample goodness-of-fit and out-of-sample forecast performance. It is calculated as the average of the unsigned errors.

$$MAD = \frac{1}{n}\sum\left|Y_t - F_t\right|$$

MAD/Mean Ratio. This MAD/Mean ratio is an alternative to the MAPE that is better suited to intermittent and low-volume data. Percentage errors cannot be calculated when the actual equals zero and can take on extreme values for low volume data. These issues become magnified when you start to average MAPEs over multiple time series. The MAD/Mean ratio tries to overcome this problem by dividing the MAD by the Mean—essentially rescaling the error to make it comparable across time series of varying scales. The statistic is calculated exactly as the name suggests—it is simply the MAD divided by the Mean.

GMRAE. The GMRAE (Geometric Mean Relative Absolute Error) is used to measure out-of-sample forecast performance. It is calculated using the relative error between the naïve model (random walk) and the currently selected model. A GMRAE of 0.54 indicates that the size of the current model's error is only 54% of the size of the error generated using the naïve model for the same data set.

Box-Cox Power Transforms

It is assumed in both Box-Jenkins and dynamic regression that the error process e_t is independently identically normally distributed. Heteroscedasticity of the error process, i.e. changing variance, is a violation of this assumption. The effect of heteroscedasticity is not so damaging as autocorrelation because it does not bias estimates of the coefficients. Its main effect is to reduce statistical efficiency, so that the effect of sampling errors is greater.

Sometimes heteroscedasticity can be eliminated or reduced by transforming the dependent variable. The transformed variable is forecasted, then back transformed to the original distribution. The most important are the power transforms analyzed by Box and Cox [1964]. The following equations are used to transform the original data Y_t to the transformed data $Y_t(\lambda)$.

$$Y_t(\lambda) = \frac{Y_t^{\lambda} - 1}{\lambda} \quad (\lambda \neq 0)$$

$$Y_t(0)=\ln(Y_t)$$

The parameter λ specifies the power to which the data are raised, except when it is zero. In that case, Y_t is replaced by its logarithm. The first of the two equations includes constant terms to make the transform a continuous function of λ.

The Box-Cox power transform can be applied only to positive data.

Safety Stocks

Forecast Pro generates safety stock calculations in addition to point forecasts and confidence limits. This capability is most often used in setting inventories which are replenished only at certain variable or fixed intervals.

Inventory control analysis requires the manager to balance inventory holding costs, reorder costs and other factors to determine economic order sizes and reorder points to maintain a desired service level at minimum cost. The analysis must take into account the lead time between placing an order and placing the units in stock. Although such analyses can become very complex, most of them require answering a question similar to the following.

How much stock do I need to maintain a service level of 95% if the reorder lead time is four weeks?

At each point of time, the manager needs enough stock so that the total sales for the next four weeks will exceed the stock level only five percent of the time. It is easy to calculate the expected demand over the four week period—just add the forecasts over the four weeks. The difficulty lies in computing the probability that sales will exceed the cumulative forecast by some certain amount. To determine this mathematically is a complex problem that depends upon the details of the statistical forecast model. Most MRP systems use a very crude approximation to solve this problem and really must, since the system does not know anything about the statistical forecast model. The difficulty of the calculation lies in taking into account the serial correlations of sales from point to point over the reorder cycle.

Forecast Pro is unique in providing a rigorous statistical solution to this problem. It does so by converting the model to an equivalent but different form called the Wold representation. This is the key to determining the statistical distribution of the cumulative forecast. Consult Wold[1938] for details of the computation.

But there is one caveat, the computation assumes that the statistical distribution of future sales is in fact correctly captured by Forecast Pro. This is never absolutely true, so the safety stocks will be in error to the extent that the Forecast Pro model does not actually capture the true model.

Consider the following example:

Date	5% Lower	Forecast	95% Upper
2009-35	640	757	875
2009-36	654	774	895
2009-37	642	761	880
2009-38	574	680	787
2009-39	572	679	786
2009-40	531	631	731

The forecast for each week represents the mean of all possible futures. Put another way, it is equally likely, according to the model, that the actual value will be above or below the forecast.

The upper and lower confidence limits provide information about the spread around the forecast for a given period. The 95 percent upper confidence limit for week 35 is 875. Thus, according to the model, actual sales for week 35 should fall at or below 875, 95% of the time. The 95 percent upper confidence limit for week 38 is 787. Thus, according to the model, actual sales for week 38 should fall at or below 787, 95% of the time.

Notice that forecasts and confidence limits do not take into account lead times. Therefore, they cannot be used to answer our question, "*How much stock do I need to maintain a service level of 95% if the reorder lead time is four weeks?*"

Lead time	DDLT	95% Safety	Reorder Point
1	757	117	875
2	1,532	174	1,706
3	2,292	215	2,507
4	**2,973**	**244**	**3,217**
5	3,652	269	3,920
6	4,283	289	4,572

The expected Demand During Lead Time (DDLT) is the cumulative forecast. Thus, for a lead time of 4 weeks, the DDLT is 2973 (757+774+761+680).

The safety stock is the excess stock needed, above and beyond the DDLT, to maintain the service level specified for the upper confidence limit percentile. The safety stocks are output for each lead time up to and including the forecast horizon. Thus, to determine the stock required for a four-week lead time, you would add together the DDLT and Safety Stock values for lead time 4 (2973+244=3217). This quantity is known as the Reorder Point. If your stock falls below the Reorder Point then you do not have enough stock to satisfy the expected demand at the specified service level and need to obtain additional stock (e.g., to reorder).

Outlier Detection and Correction

An outlier is a data point that falls outside of the expected range of the data (i.e., it is an unusually large or small data point). If you are forecasting a time series that contains an outlier there is a danger that the outlier could have a significant impact on the forecast.

One solution to this problem is to screen the historical data for outliers and replace them with more typical values prior to generating the forecasts. This process is referred to as outlier detection and correction.

Correcting for a severe outlier (or building an event model for the time series if the cause of the outlier is known) will often improve the forecast. However if the outlier is not truly severe, correcting for it may do more harm than good. When you correct an outlier, you are rewriting the history to be smoother than it actually was and this will change the forecasts and narrow the confidence limits. This will result in poor forecasts and unrealistic confidence limits when the correction was not necessary.

It is the authors' opinion that outlier correction should be performed sparingly and that detected outliers should be individually reviewed by the forecaster to determine whether a correction is appropriate.

Forecast Pro incorporates an automated algorithm to detect and (optionally) correct outliers.

The detection/correction algorithm works as follows:

1. The specified forecasting model is fit to the time series, the residuals (fitted errors) are generated and their standard deviation is calculated.

2. If the size of the largest error exceeds the outlier threshold, the point is flagged as an outlier and the historic value for the period is replaced with the fitted value.

3. The procedure is then repeated using the corrected history until either no outliers are detected or the specified maximum number of iterations is reached.

In a multiple-level problem the detection is only performed on the end items (i.e., the nongroup level). If the correction option has been selected, after all end items are corrected, the group level totals are reaggregated to reflect the corrected values.

Trading Day Effects

In many cases, the sales or shipments of a product depends upon calendar effects such as the number of working days, the number of weeks in the period (4-4-5 data), or the number of weekend days. In more complex cases, each day of the week is associated with a weight that describes its relative contribution to sales or shipments, so the weight for the month can be computed from the relative contributions of the days that fall into that particular month.

Trading day effects of this kind can influence sales by as much as five to ten percent. Since it is one variable that can be computed accurately for the future as well as the past, it makes a good deal of sense to account for it. This will usually give a boost in accuracy at very little extra effort or expense.

Forecast Pro supports a weighting transformation (\WGT=) that takes trading day effects into account in a very simple way.

> The trading day weights (both past and future), must be defined as a helper variable in a file supplied by the user. The data must span from the first historic data period to the last forecast period..
>
> The actual historic values for a time series are adjusted by dividing through by the weight for each month. This gives estimates for the sales that would have occurred in the absence of trading day effects.
>
> Forecasts are prepared from the adjusted historic data.
>
> The forecasts are multiplied by the corresponding future trading day weights.

The user must compute and supply to Forecast Pro Unlimited the appropriate trading day weights for each month (or quarter)

Methodology of Automatic Forecasting

This chapter describes special methodological considerations that apply to automatic forecasting of hundreds or thousands of items.

Introduction

Much of the forecasting literature (and much of the available software) concentrates on forecasting time series in an R&D environment. The literature envisions the forecaster as intensely interested in just one or two complex time series, perhaps for long forecast horizons, and often with highly significant consequences. For example, one might be interested in forecasting the economic or social environment for a nuclear power generation plant. The forecaster is willing, under such circumstances, to invest considerable time and other resources to obtain the best available forecasting models, and might be willing to use extremely complex methodology.

This emphasis on the R&D environment ignores the forecaster who needs to forecast hundreds or thousands of products, on a weekly, monthly or quarterly basis, perhaps for inventory control or production planning. In this case, the consequences of error in the forecasts for any particular product may be quite small, although the consequences for aggregate performance might be large. The methodology of Forecast Pro Unlimited, is directed squarely at this forecasting environment.

Forecast Pro Unlimited is based on the fairly scanty published research and upon BFS research and experience. We will summarize some of the facts that have emerged from the research.

The forecasting methods that succeed best are relatively simple ones. Product data is often so volatile that more complex models, no matter how well they fit the historical data, yield inferior forecasts.

When historic records are relatively long and not very noisy, there is a substantial payoff in matching the forecast model to each data set individually.

On the other hand, many business time series are extremely noisy. The information in an individual historic record may not be sufficient to choose the best method reliably. Frequently, method A may appear to be superior to method B at one time, and inferior to it at another. In these cases, overall accuracy may be improved by selecting a model at the group level. This can be done by using the Forecast Pro Unlimited out-of-sample evaluation procedure.

Forecasting performance and goodness-of-fit of a model to the historic data is certainly related, but the relationship is much looser than one would expect.

Forecast Pro Unlimited includes extensions (discrete distributions, Croston's intermittent data model and multiplicative error model) to the standard confidence limit methodology to make confidence limits more accurate. Nevertheless, confidence limits are only rough guides to real out-of-sample forecast performance. You can evaluate real out-of-sample performance by using the Forecast Pro Unlimited evaluation methodology.

Classification of Time Series

From the earliest days of forecasting systems for production and inventory control, it has been recognized that there are essentially three types of product time series. This stratification is based mainly on sales or production volume.

Type A series are very high volume. These series are usually fairly regular, so statistical forecasting methods like those in Forecast Pro Unlimited perform well. However, these high volume items are also of great importance to the firm, and the consequences of forecast error can be significant. Thus, if there are not too many of them, it is wise to examine them interactively, and to make judgmental adjustments as appropriate.

Type B series are of medium volume. Ordinarily, these series can be forecasted fairly accurately by the methods in Forecast Pro Unlimited. Since these items are not separately as crucial to the firm, they lend themselves well to automatic forecasting. Human intervention is usually required only when the forecasting software marks them as exceptional.

Type C series are of lowest volume, and may include as many as 50% of the total. Many of these series will be mostly zeroes, with occasional small sales and, more rarely, a large sale. The percent error of forecasts of Type C series is often quite large, but the consequence of error is usually small. When automatic forecasting first emerged, Type C series were not usually forecasted at all. Instead, a default forecast (say zero or one) was used, to save computer time (then a scarce resource). Now that computation is cheap, methods like those in Forecast Pro Unlimited are likely to provide increased accuracy.

Any of these groups can include *rogue series*, i.e., series that are so irregular as to be virtually unforecastable. Obviously, most rogue series are of Type C, so that their practical significance is usually not high. However, their influence on forecast performance evaluations can be great, since evaluations usually depend upon percent error, not absolute error. The meaningfulness of forecast performance evaluation, discussed below, will be greatly enhanced if the evaluation is based upon a classification of series, and the identification of rogue series.

Multiple-level Forecasting

Multiple-level models apply to data which must be dealt with at several levels of aggregation. Product data, for instance, often involve SKU's (stock keeping units), brands and lines. Forecast Pro Unlimited allows you to aggregate data into a hierarchy of groups, and to produce consistent forecasts at all levels of aggregation.

Consider the product group ABC consisting of the sum of products A, B and C. If one forecasts each series independently, the forecast of ABC will differ from the sum of the forecasts of products A, B and C. Often it is essential for the firm, however, to reconcile such hierarchical inconsistencies.

There are two generally accepted ways to do this—top-down and bottom-up. You will need to use your knowledge of the products, or testing, to determine which method is superior.

The top-down method forecasts ABC, A, B and C first, as a preliminary step. Then the forecasts of A, B and C are adjusted proportionately at each step of time, to insure that ABC = A+B+C. The bottom-up method is to forecast A, B and C, and to construct the forecast for ABC by summing the forecasts for A, B and C. Neither method is superior to the other under all circumstances. If the items are very similar, like sizes and colors of a product, then the top-down approach is probably more accurate. If they are disparate, like a household product and a business product, then the bottom-up method is more likely to succeed.

Many companies employ a disaggregation procedure instead of forecasting each item at the lowest level. For example, the distribution of shoe sizes and colors is relatively constant from style to style. It therefore makes sense to forecast at the style level and then to apply the overall size-color break down to obtain SKU level forecasts. The disaggregated SKU forecasts would almost certainly be more accurate than those obtained directly from SKU level data because of its extreme noisiness.

Bunn and Vassilopoulos [1993] showed that it is often more accurate to extract seasonal indexes at the group level, use them to deseasonalize the item level data, forecast the resulting nonseasonal series and reseasonalize the forecasts.

For instance, suppose that the lowest level item is a soft drink in a certain size and container type (glass, plastic, aluminum, etc.). One would extract seasonal indexes after aggregating the data over size and container type. This is especially useful if the data includes some new container sizes without adequate history to calculate seasonality independently. Another situation where this capability is important, is when products are constantly being replaced with new models. For example, the average life span of a computer printer model is 18 months, making independent estimation of the seasonal indexes very tenuous. Extracting seasonality at the group level (containing all models past and present) and applying the indexes to the current models results in more reliable estimates.

Forecast Pro Unlimited makes it easy to apply top-down seasonality. One merely marks the group with the modifier \INDEXES to compute both seasonal and event indexes at that level.

Although the above examples all use two levels of aggregation, Forecast Pro Unlimited allows you to define as many levels as you desire. Any combination of top-down and bottom-up reconciliation can be used.

Forecast Pro Unlimited disaggregates in a series of steps. First the forecasts for a top-down group are frozen. Then the first-generation component forecasts are adjusted so that they sum to their parent forecasts. Then those forecasts are frozen and used to adjust *their* component forecasts. This process continues until all item-level forecasts have been adjusted.

Incorporation of Additional Information

It will often be appropriate to adjust forecasts from Forecast Pro Unlimited judgmentally. This may be done to include information about orders that have been received, or which are expected; the effects of promotions; knowledge about external economic trends; product mix changes, etc. These are particular cases where the information available to the user is greater than that available to Forecast Pro Unlimited.

In other cases, i.e., when the additional information possessed by the manager is not clear, the effect of subjective intervention by the user may be counterproductive. Studies have shown that, although most managers *believe* that their subjective assessment is superior to quantitative projection, that is not always the case. As Fildes [1990] has put it, manager intervention is a “mixed blessing.” We advise the user to be cautious.

In any case, user intervention is time consuming. In those research studies where subjective intervention provided improved accuracy, the managers typically used graphical analysis and fairly intensive consideration as their tools. Quick “eyeball” adjustments are not likely to contribute much to accuracy.

Forecast Pro Unlimited offers a forecast adjustment facility to permit subjective intervention by the user. Both individual and group level forecasts can be adjusted in a variety of ways.

Selection of Forecasting Method

Forecast Pro Unlimited offers the user five basic forecasting methods, and a wide range of variations on these methods. The methods that are included were selected from among those few univariate forecasting models that are supported by the research, and which can be automated without excessive computational burden. Although true multivariate methods are not feasible in automatic systems, Forecast Pro Unlimited does support event modeling which can accommodate promotional schedules and business interruptions.

The five basic methods are simple moving average, exponential smoothing, Croston’s intermittent demand, discrete distributions and Box-Jenkins (ARIMA). Simple moving average is included only for use on very short data series, where it is infeasible to fit more complex models to the data. The Croston’s model is designed for data with numerous zeros. Discrete distributions (negative binomial and Poisson distributions) are for use on data whose values are small integers. The other two “methods,” exponential smoothing and ARIMA, are not actually single methods but, rather, families of methods. The member methods differ mainly in their accommodation of structural characteristics in the data like trend, seasonality and random noise. Choosing a method thus involves two steps: choosing a family and choosing a method from within the family.

Empirical research studies such as the M-Competition have shown that there is no one single forecasting methodology that is most accurate in all cases. Fildes [1990] has shown that improvements in accuracy of 20% or more can be obtained by selecting the method that is most appropriate for a given data set. Thus the task of selecting a method is crucial for accuracy.

Features such as data length, stability, trending and seasonality will often lead the experienced forecaster to favor one method over another. These factors may help the experienced forecaster form a “hunch” about the best method for a particular group of data. However, as Fildes [1990] has demonstrated, the best and most reliable way to fit the method to the data is through testing. That is why Forecast Pro Unlimited includes an expert selection algorithm to automatically select a method for each series and a facility for testing over a hold-out sample.

Thus there are three approaches to selecting a method. The first is to allow the program’s expert selection algorithm to choose the models. The second is to allow the program to select a model after you have decided on a family of models. The third is to make the selection yourself after hold-out testing over your data (preferably all of it).

Forecast Pro Unlimited’s expert selection is easy to use and works extremely well. The only disadvantages is that the algorithm is time consuming. Manual model selection allows you to consider more models during the experimentation stage. Although this step is time consuming, your forecast production runs will be substantially quicker.

The following sections describe these processes in greater depth. However, they do not cover the statistical foundations of the forecast methods themselves. They concentrate instead on implementation as automatic methods, and on implications for use of Forecast Pro Unlimited. Background material on the five methodologies (simple moving average, exponential smoothing, Croston’s intermittent demand, discrete distributions and Box-Jenkins) is presented in the next chapter.

Model Selection via Out-of-Sample Testing

Instead of letting Forecast Pro Unlimited decide which forecast model to use for each series separately, you can specify it yourself by adding the appropriate modifiers to your script file. For example, \SIMPLE forces use of simple exponential smoothing, \HOLT forces use of the Holt model, and \EXSM=LA evokes additive Winters. These three options might be appropriate for nontrended, trended and seasonal data, respectively.

The modifiers on a particular line of the script file apply to all items in the data file (or ODBC table) cited on that line. Therefore you should classify your items into different groups (and data files) with similar properties. In this way you can avoid fitting seasonal models to nonseasonal data, trended models to nontrended data, etc. Each data file is cited on a different line of the script, along with the model specification.

At first thought, this procedure seems inferior to that of selecting a model separately for each item. In fact, however, overall forecasting performance may be markedly improved. That is because business data is often so irregular that the statistical information in a single series may not be sufficient to make a reliable choice of model.

After we have classify the items into groups of like items, we must decide on a model that is best overall. This is done by using out-of-sample testing.

To test a particular model, define a holdout sample on the dialog bar before creating the forecasts. This directs Forecast Pro Unlimited to withhold the specified number of time points from the end of the data, and to fit the model to the remaining data, which we call the “fit set.” The withheld data is called the “check set.”

Forecast Pro Unlimited first forecasts the check set data from the last point of the fit set. Then it moves to the first point in the check set as a forecast base, and forecasts the remaining n-1 values. This process continues over all but the last point in the check set. Forecast errors are computed by subtracting the known true values of the check set from their forecasts.

By rolling forward in this way, Forecast Pro Unlimited accumulates a total of n one-step-ahead forecast errors from n different forecast bases, n-1 two-step-ahead errors from n-1 different forecast bases, etc. Since forecast performance can change radically from one forecast base to another, the rolling forecast errors provide a much better picture of true out-of-sample performance than a “snapshot” taken at only one forecast base.

The line-level summary statistics of the evaluation are written to the text window, where you can print it if you wish. The results from each model are summarized, so you can choose the model that performed best overall. You will probably make repeated runs, making small changes between each run. You might, for instance, reclassify your seasonal and nonseasonal series. Finally, of course, you will have the information you need to select the best methodology for your data. While this approach requires more initial work than using expert selection, it could pay off in terms of improved accuracy.

While this example focuses on different exponential smoothing models, you can also use the Forecast Pro Unlimited out-of-sample evaluation methodology to choose between automatic exponential smoothing and automatic Box-Jenkins, or to fine-tune some of the Box-Jenkins options.

Glossary

This glossary contains definitions of the technical terms used in the main body of the text. A particular definition may involve terms that are defined elsewhere in the glossary.

ACF (autocorrelation function)
The ACF consists of the autocorrelations for lags 1, 2, 3, ... N. Forecast Pro displays the ACF as a correlogram, i.e. a bar chart of the autocorrelations arranged by lag.

ARIMA model
(AutoRegressive Integrated Moving Average) model. A family of sophisticated statistical models used by Box and Jenkins to describe the autocorrelations of a time series data. The symbol ARIMA(p,d,q) indicates a model involving p autoregressive terms and q moving average terms, applied to data that have been differenced d times. The Box-Jenkins technique involves (1) Identification of a particular ARIMA model to represent historic data; (2) Estimation of ARIMA model coefficients, (3) Statistical validation of the model; and (4) Preparation of forecasts.

Autocorrelation
The correlation of a variable and itself N periods later, and hence a measure of predictability.

Base
The forecast base is the time point from which forecasts are prepared.

BIC (Bayes information criterion)

A model selection criterion proposed by Schwarz [1978]. Within a model family (e.g. exponential smoothing or Box-Jenkins), the model that minimizes the BIC is likely to provide the most accurate forecasts. Since models with many parameters often fit the historical data well, but forecast poorly, the BIC balances a reward for goodness-of-fit with a penalty for model complexity. If your current model yields the lowest BIC out of the models you have tested, Forecast Pro marks it with “Best thus far.”

Box-Cox power transform

Logarithmic or power transform of the data. Used to reduce or eliminate dependence of the local range of a time series on its local mean.

Box-Jenkins

Strictly speaking, the statistical technique developed by Box and Jenkins to fit ARIMA models to time series data. More loosely, the term refers to the ARIMA models themselves.

Confidence limits

A forecast is generally produced along with its upper and lower confidence limits. Each confidence limit is associated with a certain percentile. If the upper confidence limit is calculated for 97.5% and the lower for 2.5%, then actual values should fall above the upper confidence limit 2.5% of the time, and below the lower confidence limit 2.5% of the time. These are often called the 95% confidence limits to indicate that the actual value should fall inside the confidence band 95% of the time. In practice, confidence limits tend to overstate accuracy. You can set the confidence limit percentiles in Configure.

Dependent variable

The variable you want to forecast. Strictly speaking this term only applies to regression modeling, where there are independent variables as well, but it is sometimes convenient to use it for the variable in univariate models as well.

Differencing

To difference a time series variable is to replace each value (except for the first) by its difference from the previous value. The seasonal difference replaces each value (except for those in the first year) by its difference from the value one year previously.

Durbin-Watson test

This statistic checks for autocorrelation in the first lag of the residual errors. It should be about 2.0 for a perfect model. Forecast Pro computes the Durbin-Watson *d*-statistic, which is, strictly speaking, applicable only for regressions that include a constant intercept term, but do not include lagged dependent variables.

Exogenous variable

An exogenous variable is an explanatory variable that can be treated as a time series of ordinary numbers. Practically speaking, *independent variable* means the same thing.

Exponential smoothing — A robust forecasting method that extrapolates smoothed estimates of level, trend, and seasonality of a time series.

Fit set — The historic data set used to fit the parameters of a model, and as the base of extrapolation for the forecasts.

Forecast error — Standard error of the within-sample forecasts, computed by running the forecast model through the historic data. Used as an estimate of the one-step forecast error.

Forecast horizon — Number of periods you wish to forecast.

Forecast scenario — A forecast scenario extends the historic series of independent variables into the future. Dynamic regression forecasts are dependent on the forecast scenario.

Lag — The time difference between a time series value and a previous value from the same series.

Ljung-Box test — Checks for autocorrelation in the first several lags of the residual errors. If the Ljung-Box test is significant for a correlational model (Box-Jenkins or dynamic Regression) then the model needs improvement. The test is significant if its probability is > .99, in which case it is marked with two asterisks in the standard diagnostic output.

Local level — See local mean.

Local mean — The average level of a time series in the general neighborhood of a given point in time. Sometimes called the local level.

Local trend — The average rate of increase of a time series in the general neighborhood of a given point in time.

MAD — Mean Absolute Deviation. This measure of goodness-of-fit is calculated as the average of the absolute values of the errors. It is an important statistic in *rolling simulation* analysis.

MAPE — Mean Absolute Percentage Error. A statistic used to measure within sample goodness-of-fit and out-of-sample forecast performance. It is calculated as the average of the unsigned percentage errors.

Model — A forecasting model is an equation, or set of equations, that the forecaster uses to represent and extrapolate features in the data.

Model complexity	Model complexity is measured by the number of parameters that must be fitted to the historic data. Overfitting, i.e., using too many parameters, leads to models that forecast poorly. The BIC can help to find the model that properly trades off goodness-of-fit in the historic fitting set, and its model complexity.
Multivariate	Involving more than one variable at a time. Dynamic regression is a multivariate technique.
Residual error	The difference between a predicted value and a true value in the fitting set, i.e. the fitted error.
Robust	A robust method is insensitive to moderate deviations from the underlying statistical assumptions.
Root mean squared error (RMSE)	A statistic that is used as an indication of model fit. It is calculated by taking the square root of the average of the squared residual errors.
Seasonality	Periodic patterns of behavior of the series. For instance, retail sales exhibit seasonality of period 12 months. Usually the forecaster must take seasonality explicitly into account during the model fitting process.
Stochastic	A process is said to be stochastic when its future cannot be predicted exactly from its past. In a stochastic process, new uncertainty enters at each point in time.
Univariate	Involving only one variable at a time. Exponential smoothing and Box-Jenkins are univariate techniques.

Bibliography

J. S. Armstrong [1983] Relative Accuracy of Judgmental and Extrapolative Methods in Forecasting Annual Earnings, *Journal of Forecasting*, **2**, pp. 437-447.

J. S. Armstrong [1985] *Long Range Forecasting From Crystal Ball to Computer* (2nd ed.), New York: Wiley.

J. S. Armstrong [2001] *Principles of Forecasting: A Handbook for Researchers and Practitioners*, Norwell MA: Kluwer Academic Publishers.

R. Ashley [1983] On the Usefulness of Macroeconomic Forecasts as Inputs to Forecasting Models, *Journal of Forecasting*, **2**, pp. 211-223.

R. Ashley [1988] On the Relative Worth of Recent Macroeconomic Forecasts, *International Journal of Forecasting*, **4**, pp. 363-376.

F. Bass [2004] A New Product Growth Model for Consumer Durables, *Management Science*, **50**, pp.1825-1832.

B. L. Bowerman and R. T. O'Connell [1979] *Time Series and Forecasting: An Applied Approach*, Boston: Duxbury Press.

G. E. P. Box and G. M. Jenkins [1976] *Time Series Analysis: Forecasting and Control*, Revised Edition, San Francisco: Holden-Day.

R. G. Brown [1963] *Smoothing, Forecasting and Prediction of Discrete Time Series*, Englewood Cliffs, NJ: Prentice-Hall.

D. W. Bunn and A. I. Vassilopoulos [1993] Using Group Seasonal Indices in Multi-item Short-term Forecasting, *International Journal of Forecasting*, **9,** 4, pp. 517-526.

Caceci and Cacheris [1984] Fitting Curves to Data, *Byte Magazine*, May, 1984.

C. Chatfield [1978] The Holt-Winters Forecasting Procedure, *Applied Statistics*, **27**, pp. 264-279.

C. Chatfield and M. Yar [1988] Holt-Winters Forecasting: Some Practical Issues, *The Statistician*, **37**, pp. 129-140.

C. Chatfield and M. Yar [1991] Prediction Intervals for the Holt-Winters Forecasting Procedure, *International Journal of Forecasting*, **6**, 1, pp. 127-137.

G. C. Chow [1960] Tests of Equality between Subsets of Coefficients in Two Linear Regressions, *Econometrica*, **37**, pp. 591-605.

D. Cochrane and G. H. Orcutt [1949] Application of Least Squares Regression to Relationships Containing Autocorrelated Error Terms, *Journal of the American Statistical Association*, **44**, pp. 32-61.

J. D. Croston [1972] Forecasting and Stock Control for Intermittent Demands, *Operational Research Quarterly*, **23(3)**, pp. 289-303.

E. B. Dagum [1982] Revisions of Time Varying Seasonal Filters, *Journal of Forecasting*, **1**, 2, pp. 173-187.

D. A. Dickey, W. R. Bell, and R. B. Miller [1986] Unit Roots in Time Series Models: Tests and Implications, *The American Statistician*, **40**, pp. 12-26.

D. A. Dickey and W. A. Fuller [1979] Distribution of the Estimators for Autoregressive Time Series with a Unit Root, *Journal of the American Statistics Society*, Vol. 74, pp. 427-431.

D. A. Dickey and W. A. Fuller [1981] The Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root, *Econometrica*, **49**, pp. 1057-1072.

R. F. Engle [1984] Wold, Likelihood Ratio, and Lagrange Multiplier Tests in Econometrics, *Handbook of Econometrics Vol. II*, Eds. Z. Griliches and M.D. Intriligator, North-Holland Publishers BV.

R. Fildes [1979] Quantitative forecasting -- the state of the art: extrapolative methods, *Journal of the Operational Research Society*, **30,** pp. 691-710.

R. Fildes and S. Howell [1979] On Selecting a Forecasting Model, Forecasting, *Studies in the Management Sciences*, Vol. 12, S. Makridakis and S. C. Wheelwright (Eds.), Amsterdam: North-Holland.

A. D. Flowers and M. H. Finegan [1988] A Comparison of Microcomputer Forecasting Packages, ORSA/TIMS.

W. A. Fuller [1976] *Introduction to Statistical Time Series*, New York: Wiley.

E. S. Gardner, Jr. [1983] Automatic Monitoring of Forecast Errors, *Journal of Forecasting,* **2**, 1, pp. 1-21.

E. S. Gardner, Jr. [1985] Exponential Smoothing: The State of the Art, *Journal of Forecasting,* **4**, 1, pp. 1-38.

R. L. Goodrich [1989] Applied Statistical Forecasting, Belmont, MA: Business Forecast Systems, Inc.

C. W. J. Granger [1989] Combining Forecasts—l Twenty Years Later, *Journal of Forecasting*, **8**, pp. 167-173.

C. W. J. Granger and G. McCollister [1978] Comparison of Forecasts of Selected Series by Adaptive, Box-Jenkins, and State Space Methods, ORSA/TIMS Meeting, Los Angeles, Summer, 1978.

C. W. J. Granger and P. Newbold [1977] *Forecasting Economic Time Series*, New York: Academic Press.

G. W. C. Granger and R. Ramanathan [1984] Improved Methods of Combining Forecasts, *Journal of Forecasting*, **3**, pp. 197-204.

A. C. Harvey [1981] *Time Series Models*, Oxford, England: Phillip Allen.

G. Hill and R. Fildes [1984] The Accuracy of Extrapolation Methods; An Automatic Box-Jenkins Package Sift, *Journal of Forecasting*, **3**, pp. 319-323.

C. C. Holt [1957] Forecasting trends and seasonals by exponentially weighted moving averages, O .N. R. Memorandum, No. 52, Carnegie Institute of Technology (now Carnegie-Mellon University).

K. Kahn [2006] New Product Forecasting: An Applied Approach, Armonk NY: M.E. Sharpe.

A. B. Koehler and B. L. Bowerman [1989] Prediction Intervals for ARIMA Models, Ninth International Symposium on Forecasting, Vancouver.

A. B. Koehler and E. S. Murphree [1986] A Comparison of the AIC and BIC on Empirical Data, Sixth International Symposium on Forecasting, Paris.

R. Lewandowski [1982] Sales Forecasting by FORSYS, *Journal of Forecasting*, **1**, pp. 205-214.

J. Ledolter and B. Abraham [1984], Some Comments on the Initialization of Exponential Smoothing, *Journal of Forecasting*, **3**, pp. 79-84.

G. Libert [1984] The M-Competition with a Fully Automatic Box-Jenkins Procedure, *Journal of Forecasting*, **3**, pp. 325-328.

G.L. Lilien, A. Rangaswamy and C. Van den Bulte [1999] Diffusion Models: Managerial Applications and Software, *Institute for the Study of Business Markets*, **7**, University Park, PA.

G. M. Ljung and G. E. P. Box [1978] On a Measure of Lack of Fit in Time Series Models, *Biometrika*, **65**, pp. 297-303.

E. J. Lusk and J. S. Neves [1984] A Comparative ARIMA Analysis of the 111 Series of the Makridakis Competition, *Journal of Forecasting*, **3**, pp. 329-332.

S. Makridakis and M. Hibon [1979] Accuracy of Forecasting: An Empirical Investigation (with discussion), *Journal of the Royal Statistical Society*, **A 142**, Part 2, 97-145. Reprinted in Makridakis et al. [1982].

S. Makridakis and M. Hibon [1989] Exponential Smoothing: The Effect of Initial Values and Loss Functions on Post-sample Forecasting Accuracy, Ninth International Symposium on Forecasting, Vancouver.

S. Makridakis and M. Hibon [2000] The M3-Competition: Results, Conclusions and Implications, *International Journal of Forecasting*, **16**, 4, pp. 451-476.

S. Makridakis and S. C. Wheelwright [1979] *Interactive Forecasting*, Second Edition, San Francisco: Holden-Day.

S. Makridakis, S. C. Wheelwright and R.J. Hyndman [1998] *Forecasting Methods and Applications*, Third Edition, New York: Wiley.

S. Makridakis et al. [1982] The Accuracy of Extrapolation (Time Series) Methods: Results of a Forecasting Competition, *Journal of Forecasting*, **2**, pp. 111-153.

S. Makridakis et al. [1984] *The Forecasting Accuracy of Major Time Series Methods*, Chichister: Wiley.

E. McKenzie [1986] Error Analysis for Winters' Additive Seasonal Forecasting System, *International Journal of Forecasting*, **2**, pp. 373-382.

D. C. Montgomery and L. A. Johnson [1976] *Forecasting and Time Series Analysis*, New York: McGraw-Hill.

J. S. Morrison [1995] Life-Cycle Approach to New Product Forecasting, *The Journal of Business Forecasting*, **14**, 2, pp. 3-5.

H. L. Nelson and C. W. J. Granger [1979] Experience with Using the Box-Cox Transformation when Forecasting Economic Time Series, *Journal of Econometrics*, **10**, pp. 57-69.

C. R. Nelson and H. Kang [1981] Spurious Periodicity in Inappropriately Detrended Time Series, *Econometrica*, **49**, pp. 741-751.

C. R. Nelson and H. Kang [1983] Pitfalls in the Use of Time as an Explanatory Variable in Regression, *Journal of Business and Economic Statistics*, **2**, pp. 73-82.

P. Newbold [1983] ARIMA Model Building and the Time Series Analysis Approach to Forecasting, *Journal of Forecasting*, **2**, 1, pp. 23-35.

P. Newbold and T. Bos [1990] *Introductory Business Forecasting*, Cincinnati: South-Western.

K. Ord and R. Fildes [2013] *Principles of Business Forecasting*, Cengage.

A. Pankratz and U. Dudley [1987] Forecasts of Power-transformed Series, *Journal of Forecasting*, **6**, pp. 239-248.

A. Pankratz [1991] *Forecasting with Dynamic Regression Models*, New York: Wiley.

R. Ramanathan [1992] *Introductory Econometrics with Applications*, Second Edition, Fort Worth: Harcourt Brace Jovanovich College Publishers.

G. Schwarz [1978] Estimating the Dimension of a Model, *Ann. Statist*. **6**, 2, pp. 461-464.

J. Shiskin, A. Young, and J. Musgrave [1967] The X-11 Variant of the Census Method II Seasonal Adjustment Program, Technical Paper No. 15, Bureau of the Census.

E. Stellwagen [2012] Exponential Smoothing: The Workhorse of Business Forecasting, *Foresight: the International Journal of Applied Forecasting*. **27**, pp. 23-28.

P. A. Texter and J. K. Ord [1989] Forecasting Using Automatic Identification Procedures, *International Journal of Forecasting*, **5**, 2, pp. 209-215.

H. Theil [1971] *Principles of Econometrics*, New York: Wiley.

N. T. Thomopoulos [1980] *Applied Forecasting Methods*, Englewood Cliffs, NJ.: Prentice-Hall.

S. C. Wheelwright and S. Makridakis [1985] *Forecasting Methods for Management*, New York: Wiley.

T. R. Willemain, C. N. Smart, J. H. Shockor and P. A. DeSautels [1994] Forecasting Intermittent Demand in Manufacturing: a Comparative Evaluation of Croston's Method, *International Journal of Forecasting*, **10**, 4, pp.529-538.

G. T. Wilson [1979] Some Efficient Computational Procedures for High Order ARMA Models, *J. Statist. Comput. Simul*., **8**, pp. 301-309.

P. R. Winters [1960] Forecasting Sales by Exponentially Weighted Moving Averages, *Management Sci*., **6**, pp. 324.

H. O. Wold [1938] *A Study in the Analysis of Stationary Time Series*, Uppsala: Almquist and Wicksell.

M. Yar and C. Chatfield [1991] Prediction Intervals for Multiplicative Holt-Winters, *International Journal of Forecasting*, **7**, 1, pp. 31-37.

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

Index

Adjusted R-square, 34
ARIMA
  processes, 25
Autocorrelation, 49
  function, 22, 27
Autoregressive
  order, 27
  processes, 25
  terms, 23
Backward shift operator, 25
Bayes information criterion, 50
Bayesian information criterion, 35
BIC, 8, 35
Box-Cox power transform, 50
Box-Cox power transforms, 36
Box-Jenkins, 20
  identification, 27
  initialization, 21
  mixed processes, 28
  parameter estimation, 21
  seasonal models, 26
Check set, 46
Cochrane-Orcutt, 30, 31
Confidence limits, 8, 50
Croston’s model, 18
Curve Fitting, 19
Dependent variable, 50
Differencing, 24, 27, 50
Durbin-Watson, 34
Durbin-Watson test, 50
Dynamic regression, 28
  diagnostics, 30
  dynamics, 31
  variable specification, 31
Event effects, 4
Event indexes, 4
Exogenous variable, 50
Exponential smoothing
  description, 3
  Holt, 6, 13
  simple, 5
  Winters, 6, 14
Fit set, 46, 51
Forecast error, 35
GMRAE, 36
Holdout sample, 46
Judgmental forecasting, 44
Lagged dependent variables, 29
Level, 4
Ljung-Box, 35
Ljung-Box test, 51
MAD, 36, 51
MAD/Mean Ratio, 36
MAPE, 36, 51
M-Competition, 45
Mean, 34
Model complexity, 52
Model selection, 8
Moving average
  order, 28
  processes, 25
  terms, 23

Multiple-level forecasting, 43
Multivariate, 52
NA-constant level model, 15
Negative binomial distribution, 17
Number of parameters, 34
Outlier detection, 39
Parameter optimization, 8
Partial autocorrelation function, 27
Poisson distribution, 16
Random events, 4
Random shocks. *See* Random events
Residual error, 52
RMSE, 36
R-square, 34
Safety stocks, 37
Sample size, 34
Seasonal effects, 4
Seasonal indexes, 4
Seasonal simplification, 15
Seasonality
    additive, 6
    multiplicative, 6
Selecting a method, 44
SMAPE, 36
Standard deviation, 34
Standard forecast error, 35
Stationarity, 21, 24
Stochastic, 52
Time series
    classification of, 42
    hierarchies, 43
    rogue, 42
    types A,B,C,, 42
Trading day effects, 39
Trend, 4
Univariate, 52