AIRPORTS COUNCIL INTERNATIONAL

ACI Guide to World Airport TRAFFIC FORECASTS 2016

AIRPORTS COUNCIL INTERNATIONAL

Contents

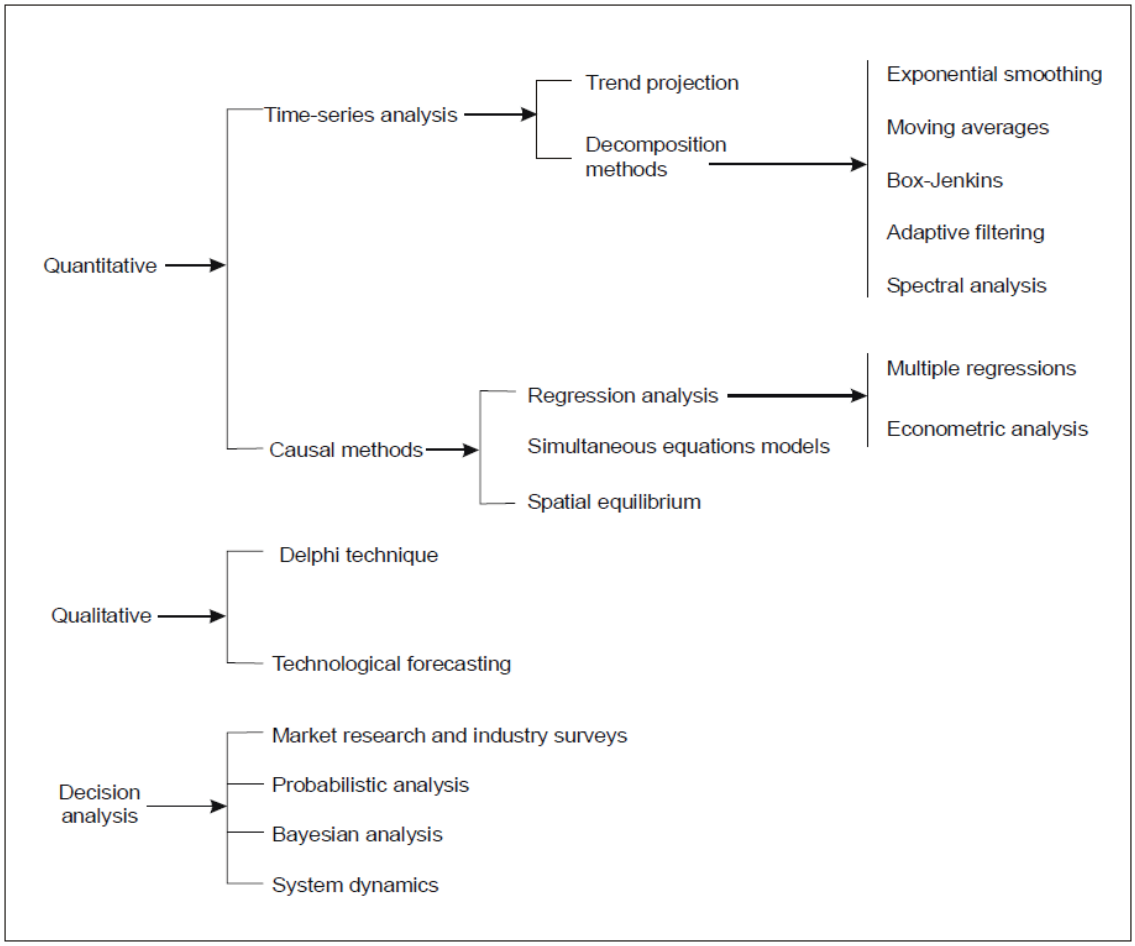
AIRPORTS COUNCIL INTERNATIONAL

Introduction

The purpose of this guide is to offer a brief, accessible review of the forecasting methods used by Airports Council International to produce aggregate airport traffic projections at the global, regional and country levels based on internationally comparable airport traffic data.[1] The guide is not intended to be an exhaustive list of forecasting techniques or an advanced statistics manual. That is, the methodologies contained herein are primarily a function of the available data and time series at the international level.

Figure 1 comprehensively depicts the major existing categories of methods to generate predictions. This guide's focus is on quantitative forecasts that are generated primarily by ACI traffic data while relying on outside sources for macroeconomic and demographic statistics.

Figure 1: Forecasting techniques



Source: International Civil Aviation Organization (ICAO) Document #8991 (2006)

[1] This guide was authored by Aram Karagueuzian, Guillaume Rodier and Daniel Sallier with contributions by Patrick Lucas, as well as valuable inputs from members of the Airport Traffic Think Tank *(at3)* peer review group on earlier drafts.

What is forecasting?

The famous economist and diplomat John Kenneth Galbraith once stated that "There are two kinds of forecasters: those who don't know and those who don't know they don't know." His observation is probably not so much a critique of the forecasting profession but rather describes the inherent uncertainty in making predictions about the future. Measuring risk typically means gauging the parameters of a *calculable* unpredictability, such as in the case of a coin toss. The probability distribution for a coin toss can be easily computed, but that in no way means the outcome can be reliably predicted. Nobody knows for sure what the result will be, but everyone agrees it'll be heads or tails—and that these alternatives are equally likely. Uncertainty refers to an *incalculable* unpredictability, like casting a die without knowing how many faces it has, what numbers are on those faces and if the die is balanced or not. Forecasting, therefore, is not risky; it's uncertain because no one knows the full list of events that could occur in the future, or how probable each one really is.

Forecasting typically implies the assumption that uncertainty can be correctly modelled as risk. There exists an underlying, "true" model that dictates the evolution of the variable of interest. Whether there really is such a model is unclear. What really matters are the implications of this assumption: that the future can be reliably predicted by analyzing the past. The variable moves in a distinguishable way; it evolves cyclically, follows a trend and exhibits seasonality. It remains unknown what will happen in the future but it is reasonable to assume that without any large disturbance, the variable will continue to evolve in a similar pattern.

Predictions are not made blindly or arbitrarily; they are made methodically from meticulous consideration of past and present observations. A forecast is a *function* of the currently available data.

$$f(x) = y$$

If the above equation is used as a representation of forecasting, then the input (x) represents available data, the function (f) represents the method used and the output (y) represents the forecasted value. Most quantitative models can be (loosely) represented in this way. Some also use other variables—and predictions of these—as input to produce forecasts. This approach is mostly used for medium- to long-term predictions, since univariate models perform quite well in the short run.

Why forecast?

Forecasts are a crucial ingredient in airport planning for the determination of future capacity requirements. Because infrastructure projects are costly and involve many resources, a data-driven understanding of future demand such as the expected number of aircraft movements, passenger traffic throughput and air cargo volumes gives airport planners and investors the necessary information for effective decision

making. Applications for these forecasts may include managing expected peak demand on both the airside and landside of an airport under a short time horizon over a span of months. On the other hand, long-term forecasts are used to plan over decades. Thus, irrespective of the element of uncertainty with future outcomes and events, forecasts are still required to understand various scenarios, all other things equal.

Forecasting horizons

Table 1 summarizes the differences in perspective between the main actors of the industry. They are compared through four forecasting horizons: very short term, short term, medium term and long term. Note that while the actual time intervals associated with these terms vary from one actor to the next, the typical purpose of a given horizon is largely invariant.

Table 1: Forecasting terms by industry

	Very Short	Short	Medium	Long
	Mostly for operational reasons and resource allocation	Operational and budget reasons	Budget reasons and investment adjustements	Strategic decisions, long term investment, product infrastructure developments
Airlines	Next flight	Current IATA season	Next 12 months	3 to 5 years
Airports	Next day to current IATA season	Current and next year	Next 5 years	Up to 20-25 years
Aircraft/engine manufacturers		Current year	Next 5 years	Up to 20-25 years
Civil aviation authorities		Current year	Next 5 years	Up to 30-40 years

When compared in measurable time, it is clear that airlines are different from other industries in the air transport value chain. The airlines' long run is roughly equal to the other industry stakeholders' medium run. This is the consequence of airlines having a level of flexibility (in implementing changes to their products, networks and fleets) that the rest of the industry's players do not. While most of the forecasting methods presented here are used in a plethora of fields and situations, this guide focuses on short-, medium- and long-term forecasting using internationally comparable airport data.

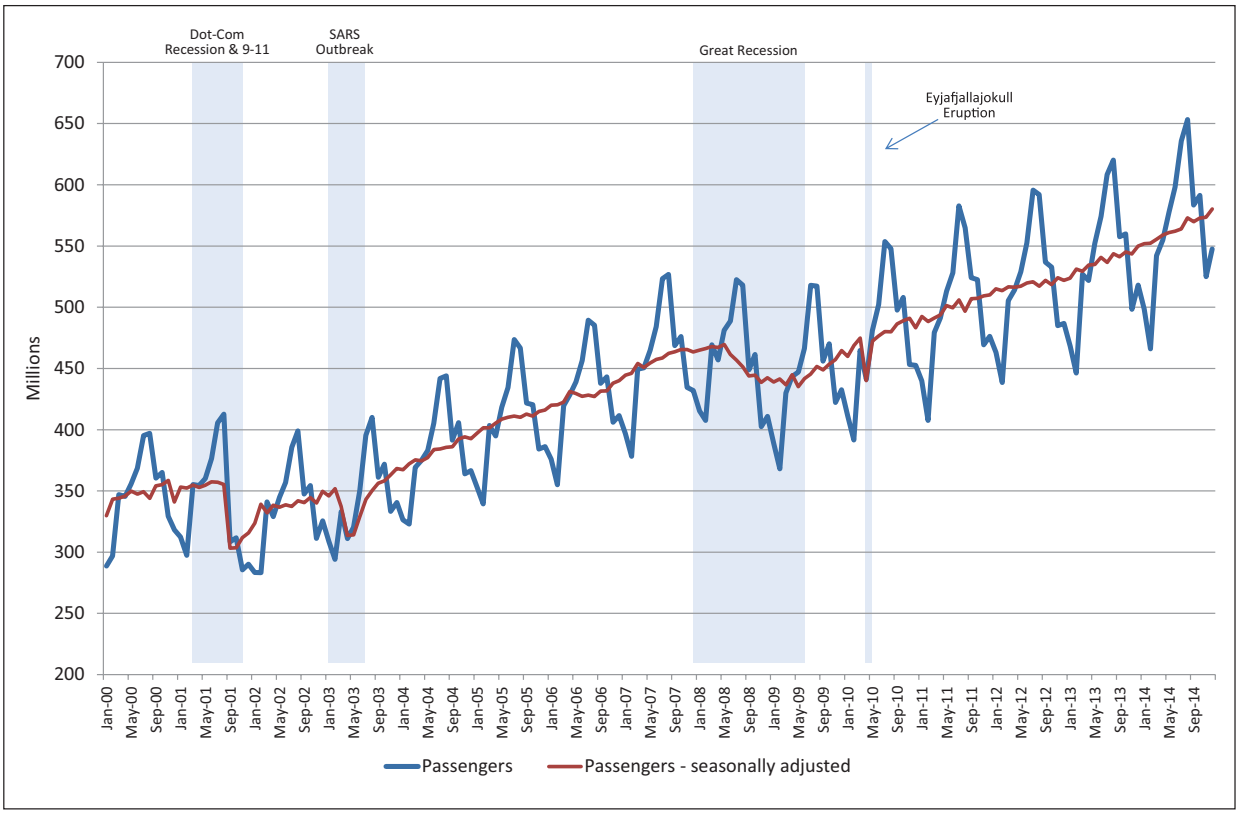
AIRPORTS COUNCIL INTERNATIONAL

Tools and techniques for short-term forecasting

ACI traffic data

Demand for air transport is subject to variations in any given year and beyond. When variability is a recurring phenomenon, the data series that describes traffic over time is said to have a seasonal component. This seasonal component is inherently non-stationary in that the behaviour of the data is dependent on time. Similarly, over longer periods of time, a cyclical component is observed, which captures either expansions or contractions in traffic. The cyclical component tends to coincide with the global business cycle. Figure 2 illustrates passenger traffic over time. An overall upward trend is observed in both the original and deseasonalized series. Random events or irregular episodes are also depicted in Figure 2.

Figure 2: Global passenger traffic (2000–2014)



Source: World Airport Traffic Database (2015)

Generally speaking, time series are sequences of measurements of a variable taken at equally spaced time intervals. The frequency at which the data is collected may be quarterly, monthly, weekly, daily and so on. Time series data consists of four components:

- Trend: long-term upward or downward movement observed in the data over several years;
- Seasonal: short-term patterns in the data that repeat themselves;
- Cyclical: a sequence of smooth fluctuations—longer than a year—around the long-term trend characterized by alternating periods of expansion and contraction; and
- Irregular: comprises the residual, erratic fluctuations of the series, which cannot be attributed to the first three or systematic components.

The trend and cycle are combined since many series available are relatively short.

Decomposition models

Decomposition techniques develop forecasting models in which time series components—trend, seasonal, cyclical and irregular—are isolated and measured. There are two types of models:

- Additive: assumes that the components of the series behave independently of each other. An increase in the cycle-trend will not cause an increase in the seasonal component; and
- Multiplicative: assumes that the components are interdependent, meaning that an increase in the trend-cycle causes an increase in the magnitude of seasonality.

Additive: $Y_t = T_t + S_t + I_t$ and $SA_t = T_t + I_t$

Multiplicative: $Y_t = T_t \times S_t \times I_t$ and $SA_t = T_t \times I_t$

In the above two equations Y_t is the original series, T_t is the trend-cycle, S_t is the seasonal component, I_t is the irregular part and SA_t is the seasonally adjusted series.

The multiplicative decomposition model uses a method called "ratio-to-moving averages" to calculate the seasonal variation, and trend analysis (using regression) to measure the trend component. The effects of the seasonal component are measured in the form of indices while the effects of the trend component are measured directly from the data.

The first step in applying this decomposition technique is to remove the trend component from the data. Several methods can be used such as linear regression, differencing or exponential smoothing. Ratio-to-moving averages is the most common method to remove the trend component. As can be surmised from its name, the method uses moving averages and subsequently ratios to develop seasonal indexes. Since the frequency of the data is monthly, a centered twelve-month moving average (or twelve-month optimally weighted moving average) for the data is calculated. The centered moving average (CMA) measures any trend effect.

The division of the trend component by its corresponding CMA would isolate the seasonal-irregular component, also known as the de-trended data.

$$SI = \frac{Y_t}{T_t} = \frac{T_t \times S_t \times I_t}{T_t} = S_t \times I_t$$

Each value of *SI* is an estimate of its corresponding seasonal index. One overall seasonal index for each month is calculated by taking the average of all matching seasonal index estimates. The resulting seasonal indices should then be normalized (the sum of the indices should add exactly to twelve) in order to obtain S_t.

These seasonal indices predict the seasonal influence of that month of the year in comparison to expected values for that segment of the year. A seasonal index greater than one implies higher values than expected for that month, and conversely, a seasonal index less than one implies smaller than expected values for that month.

The actual data Y_t is deseasonalized by dividing the monthly values by their appropriate seasonal indices.

$$SA_t = TI = \frac{Y_t}{S_t} = \frac{T_t \times S_t \times I_t}{S_t} = T_t \times I_t$$

The deseasonalized data contains only the trend and irregular components; hence linear regression is applied to estimate the inclination from the deseasonalized data. This trend equation is used to generate projections for future periods. The final forecast for each period is calculated by multiplying the forecast trend by the corresponding seasonal index.

In the same way, the additive decomposition method uses a centered twelve-month CMA to estimate the trend-cycle component which is then subtracted from the original series to obtain the seasonal-irregular component.

$$SI = Y_t - T_t = T_t + S_t + I_t - T_t = S_t + I_t$$

The seasonally adjusted series are calculated by subtracting the normalized seasonal indices from the original series.

$$SA_t = TI = Y_t - S_t = T_t + S_t + I_t - S_t = T_t + I_t$$

The forecasts are calculated by adding back the seasonal indices on the forecasted trend.

Exponential smoothing with trend and seasonal component

Four categories of triple exponential smoothing models are identified in this section. The difference lies within the types of trend and seasonal components:

- Additive trend, additive seasonality;
- Additive trend, multiplicative seasonality;
- Multiplicative trend, additive seasonality; and
- Multiplicative trend, multiplicative seasonality.

The following are the simplified sets of equations for each of the four exponential smoothing models:

$$L_t = \alpha U_t + (1 - \alpha)V_t \quad \text{Level (overall) smoothing}$$

$$b_t = \beta W_t + (1 - \beta)b_{t-1} \quad \text{Trend smoothing}$$

$$S_t = \gamma X_t + (1 - \gamma)S_{t-s} \quad \text{Seasonal smoothing}$$

Where $U_t, V_t, W_t, \text{ and } X_t$ vary in each model and are conditionally defined in table 2.

Table 2: Exponential smoothing variations

	Additive seasonality	Multiplicative seasonality
Additive Trend	$U_t = Y_t - S_{t-s}$ $V_t = L_{t-1} + b_{t-1}$ $W_t = L_t - L_{t-1}$ $X_t = Y_t - L_t$ $F_{t+m} = L_t + mb_t + S_{t+m-s}$	$U_t = Y_t / S_{t-s}$ $V_t = L_{t-1} + b_{t-1}$ $W_t = L_t - L_{t-1}$ $X_t = Y_t / L_t$ $F_{t+m} = (L_t + mb_t) \times S_{t+m-s}$
Multiplicative Trend	$U_t = Y_t - S_{t-s}$ $V_t = L_{t-1} \times b_{t-1}$ $W_t = L_t / L_{t-1}$ $X_t = Y_t - L_t$ $F_{t+m} = L_t \times b_t^m + S_{t+m-s}$	$U_t = Y_t / S_{t-s}$ $V_t = L_{t-1} \times b_{t-1}$ $W_t = L_t / L_{t-1}$ $X_t = Y_t / L_t$ $F_{t+m} = L_t + b_t^m \times S_{t+m-s}$

Note that Y_t is the observation at time t, L_t is the smoothed observation at t, b_t is the trend factor at t, S_t is the seasonal index at t and F_{t+m} is the forecast for period $t+m$ given knowledge of data up to period t. The parameters α, β and γ are constants that are generated to maximize predicted forecast accuracy.

The level L_s is initialized by calculating the average of the actual values for the first season

$$L_s = \frac{\sum_{i=1}^{s} Y_i}{s}$$

The initial b_s is assumed to be 0 or 1 for the additive or multiplicative trend respectively.

$$\begin{cases} b_s = 0 \; for \; add \; trend \\ b_s = 1 \; for \; mul \; trend \end{cases}$$

The seasonal indices are initialized by taking the data-to-average ratio of the first year.

$$S_i = \frac{Y_i}{L_s}$$

Autoregressive Integrated Moving Average (ARIMA) models

To properly grasp what ARIMA modelling entails, it is essential to understand what autoregressive and moving-average models are, as well as what differencing entails.

Autoregressive models can be written as:

$$Y_t = b_0 + b_1 Y_{t-1} + \cdots + b_p Y_{t-p} + \varepsilon_t$$

where Y_t is the value for time t, $b_0, \ldots, b_p$ are the model's parameters, $Y_{t-1}, \ldots, Y_{t-p}$ are the p last values of Y and ε_t is the current error term. This model is called an autoregressive model of order p or AR(p) in short. The value of Y is estimated at time t by adding the constant b_0 to the weighted sum of the p past observations. For example, an AR(1) would yield

$$Y_t = b_0 + b_1 Y_{t-1} + \varepsilon_t \text{ as an estimate of } Y \text{ for time } t.$$

Moving-average models can be described as:

$$Y_t = b_0 + \varepsilon_t + b_1 \varepsilon_{t-1} + \cdots + b_q \varepsilon_{t-q}$$

where Y_t is the value for time t, $b_0, \ldots, b_q$ are the model's parameters and $\varepsilon_t, \ldots, \varepsilon_{t-q}$ are the error terms of the q past periods. This model is named a moving-average model of order q or MA(q). The value of Y is estimated at time t by adding together the constant

b_0—the series' mean—to the weighted sum of the q most recent observation's errors. For example, an MA(1) would produce:

$$Y_t = b_0 + \varepsilon_t + b_1\varepsilon_{t-1} \text{ as an estimate of the value of } Y \text{ at time } t.$$

Differencing is done to remove trend and seasonality from the data. Stationarity, the property by which the distribution (mean, variance, etc.) of the data remains stable over time, is an important condition for ARIMA modeling. If the data is steadily rising, then taking the difference between the current observation and that of the previous period ($Y_t - Y_{t-1}$) is a good way to get a trendless series. This is called the first difference. If the trend is not constant but evolving at varying rates, it might be necessary to take additional differences. Seasonal differencing ($Y_t - Y_{t-12}$ for example) might also be considered if the data features seasonality. For example, if the original series (Y_t) exhibits a constant trend in addition to monthly seasonality, the modified series

$$(Y_t - Y_{t-1}) - (Y_{t-12} - Y_{t-13})$$

will be stationary, since it features first differencing and seasonal differencing of order 12. The presence of trend and seasonality is not always obvious to the naked eye. A more reliable way to detect patterns in the data is to compute its autocorrelation and partial autocorrelation functions (ACF and PACF). Autocorrelation refers to the correlation of a variable's value at a certain time (Y_t, for example) with past values of itself (Y_{t-1},Y_{t-2},…). ACF graphs the average observed correlation for each lag.

Suppose that Y_t is correlated with its lag-1 value, Y_{t-1}, with a coefficient of 0.95. If Y_{t-1} is also highly correlated to its own lag-1 value, Y_{t-2}, then Y_t will be correlated to Y_{t-2}. The autocorrelation function of Y_t will show high values at lags 1 *and* 2. It reveals that Y_t is strongly correlated with its past values, *for all t*. It can't help identify what lags truly help to determine the value of Y_t. Does the original correlation emanate from Y_{t-1} or Y_{t-2}? To answer this question, the PACF, which isolates the effect of each lag, must be used. A significant partial autocorrelation between an observation and its lag-k value means that there is a correlation between those two quantities *net of the effect of all other lags (1,2,…,k-1)*.

Table 3: Classic ACF and PACF patterns

	ACF	PACF
$AR(p)$	Exponential decay or damped sine-wave pattern	Significant spikes at lags 1 to p; then becomes much smaller
$MA(q)$	Significant spikes at lags 1 to q; then becomes much smaller	Exponential decay or damped sine-wave pattern

Table 3 shows how to interpret classic ACF and PACF figures. An exponentially decaying ACF accompanied by a PACF with significant spikes up until a certain point (p) is a good indicator that the series follows an AR process of order p. Similarly, an ACF with significant spikes from 1 to a given point (q) and a PACF that is converging towards 0 points towards an $MA(q)$ model. Some time series exhibit features of both AR and MA processes; for estimating these, a more complex (and more general) model is required.

ARIMA models, also called Box-Jenkins models, allow for the inclusion of autoregressive terms, moving average terms and differencing operations. Such models are typically noted as $ARIMA(p,d,q)$ where p is the order of the autoregressive part of the model, d is the number of differences performed and q is the order of the moving-average section of the model. For seasonal data, the notation $ARIMA(p,d,q)(P,D,Q)_{12}$ is used, where p,d,q have the same definition as before and P is the order of the seasonal autoregressive portion of the model, D is the number of seasonal differences and Q is the order of the seasonal moving-average part of the model. The form of seasonality, which here is a 12-period (month) cycle, is also indicated. The general $ARIMA(p,d,q)(P,D,Q)_{12}$ equation is

$$\left(1-\sum_{i=0}^{p}\phi_i B^{1+i}\right)\left(1-\sum_{i=0}^{P}\Phi_i B^{12+i}\right)\left(1-\sum_{i=0}^{d}B^{1+i}\right)\left(1-\sum_{i=0}^{D}B^{12+i}\right)Y_t=\left(1+\sum_{i=0}^{q}\theta_i B^{1+i}\right)\left(1+\sum_{i=0}^{Q}\Theta_i B^{12+i}\right)\varepsilon_t$$

where Y_t is the value of Y for time t and ε_t is the error at time t. The model parameters ϕ,Φ,θ,Θ are associated with standard AR, standard MA, seasonal AR and seasonal MA, respectively. The quantities p,d,q,P,D,Q have the same definition as above. The B^j are backshift operators[2] of order j. For example,

$$(1-\phi B^1)(1-\Phi B^{12})(1-B^1)(1-B^{12})Y_t=(1+\theta B^1)(1+\Theta B^{12})\varepsilon_t$$

is the formulation of an ARIMA$(1,1,1)(1,1,1)_{12}$.

ARIMA processes are a class of linear models capable of representing stationary as well as non-stationary time series by relying on autocorrelation patterns in the data. They do not involve independent variables in their construction and make use of the information in the series itself to generate forecasts.

[2] This operator is defined so that $(B^j)Y_t = Y_{t-j}$, meaning that a value Y at time t modified by the operator of order j will be equal to the value j periods ago. For example, $(B^2)Y_3 = Y_1$. The value for the third period modified by a backshift operator of order 2 is equal to the value in period 1.

Forecasting methodology with ARIMA is different from most methods since it does not assume any particular pattern in the historical data of the series to be forecasted. It uses an interactive approach of identifying a possible model from a general class of models, which is then checked against historical data to see if it accurately describes the series.

The Box-Jenkins methodology refers to a set of procedures for identifying, fitting and checking ARIMA models with time series data. If the analyst encounters a problem at any of these stages, corrections must be made. Eventually, a specified model should fill all the required criteria. Forecasts follow directly from the form of the fitted model.

The choice of p, d, q, P, D and Q can also be automated. This usually involves having an algorithm that isolates values of the model's parameters which allow for a minimal level of a criterion such as Akaike's Information Criterion (AIC). Since the chosen criterion is an indicator of a model's average forecast error, the algorithm finds the model which presumably offers the best forecast. Regardless of the method used to discriminate between models, it should involve cross-validation rules or test statistics. These are discussed in more detail later in the section on model selection.

Tools and techniques for medium- and long-term forecasts

When analyzing medium- or long-term trends, it is common to look at annual data. This means that the seasonal component of time series will be removed. The cycle and trend are the factors of interest. This new focus and time horizon implies that a different set of tools should be used.

Univariate models

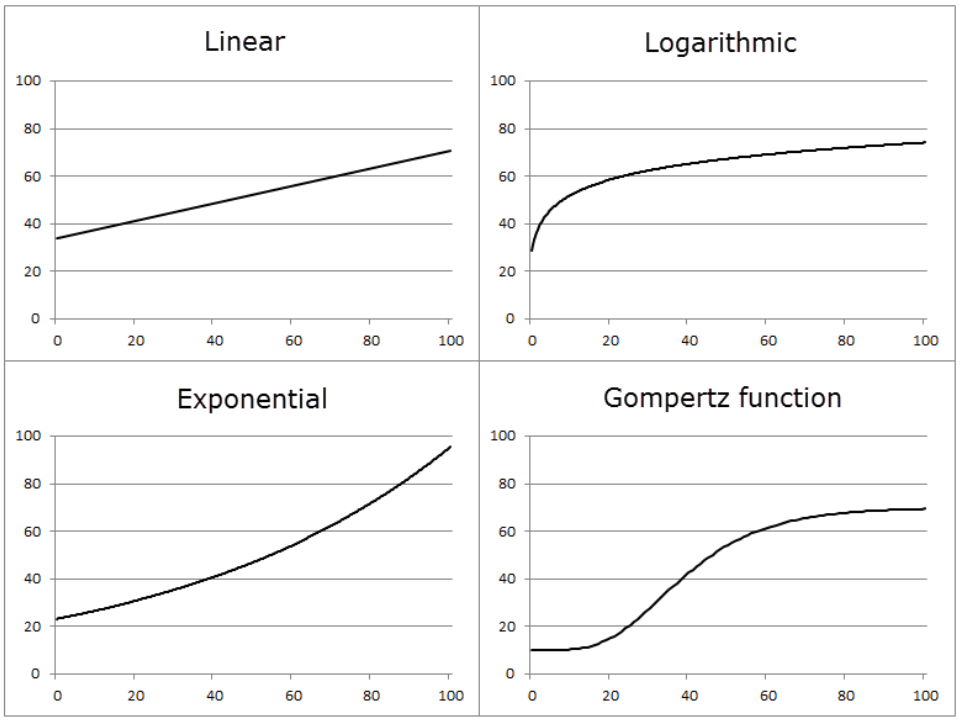
Another approach is to predict future movements of the variable of interest by analyzing the available data. While some of the methods presented in the short-term forecasting section can be modified to tend to long-term predictions, this usually involves another layer of complexity for relatively small, if any, gain in comparative forecasting accuracy.

Trend projection

A different way to generate forecasts based on the observations is to find a function that mimics its pattern in a coherent way. Practically, trend analysis usually takes the form of a function with time as a variable. To perform this, one can use annual data or 12 sliding months' cumulated data. The latter involves using monthly data and creating pseudo-yearly data. For example, suppose five years of monthly data, meaning that $5 \times 12 = 60$ data points are available. The first pseudo-yearly point will be the sum of months 1 through 12. The second will be the sum of months 2 to 13, and so on. A time series of 49 entries would be obtained where each entry y_t would be equal to the sum of the observed data for months t to month $t + 11$ (with t variying from 1 to 49).

Many functions can be written to represent the relationship between the variable of interest and time. The choice of functional form is left to the analyst's judgement. Theoretical models and past data observation can both be relevant in making this decision. Table 4 shows a few example functions that could be used to model long-term trends.

Figure 3: Examples of classic trend projections



The first is a linear trend which embodies the idea that the variable will continue to grow at a fixed pace (a). This can be approximately true for relatively short time horizons even if the general trend is rather curved. If diminishing returns are expected, a logarithmic trend might be preferred. If increasing returns are more coherent with the available data, then an exponential formulation could be picked. The Gompertz function can help model cases where a period of increasing returns is followed by a time of diminishing returns; this is a typical assumption for very long-term market trends. Other S-curves like logistic functions can serve this purpose as well. The equations for the graphed examples are as follows.

$$\textit{Linear} \qquad Y(t) = at + b$$

$$\textit{Logarithmic} \qquad Y(t) = a \times ln(t) + b$$

$$\textit{Exponential} \qquad Y(t) = a(1+b)^{t}$$

$$\textit{Gompertz} \qquad Y(t) = ae^{-be^{-ct}}$$

In all cases, $Y(t)$ is the variable to be forecasted, a, b and c are constants chosen to minimize the function's fitting errors and t is time, usually measured in periods. This is in no way an exhaustive list. Any mathematical function whose shape and properties are coherent with the data at hand and the intuition surrounding the variable's evolution is a potential candidate.

If future observations fall near the projected trend, it is safe to say that the functional form was properly chosen. Since there is no way to know if this will be the case beforehand, typical ways to choose a model function include:

- Looking for a function which mimics the properties predicted by a theoretical model of $Y(t)$;
- Trying several functional forms and picking the one which has the best fit[3]; and
- Looking at the data of related variables or similar but better documented variables to analyze their trend.

Causal models

The methods presented in the short-term forecasting section share a common characteristic: they are all univariate. In other words, they are methods in which the input is the past values of Y and the output is a function of these observations. The models predict the variable's future position by identifying patterns in its observed movements. Another fundamentally different way to generate forecasts is to identify the variables that determine Y and use this information to build a predictive model. When dealing with economic data, this approach is referred to as econometrics.[4]

Regression models

The most common form of applied econometrics is the linear regression model. It consists of an equation describing the variable of interest as a sum of a number of explanatory variables, each multiplied by a constant parameter. The general expression for a linear regression is:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + e$$

where Y is the variable of interest, $\beta_0, \ldots, \beta_k$ are the model's parameters, $x_1, \ldots, x_k$ are the variable used in the estimation and e is the error term. Sometimes the distinction is made between simple regression analysis—the case with one explanatory variable—and multiple regression analysis, which is the general case. When studying time series data, the general equation can be written as:

$$Y_t = \beta_0 + \beta_1 x_{1,t} + \beta_2 x_{2,t} + \cdots + \beta_k x_{k,t} + e_t$$

[3] Beware high degree polynomial. Typically, as the degree of a polynomial equation rises above 2 or 3, fit measures grow better but forecasts worsen. Additionally, this procedure binds the analyst to the assumption that fit accuracy and forecast accuracy share a monotonic relation.

[4] For the purpose of this guide, econometrics is defined as the use of quantitative techniques on economic data aiming to empirically study economic relationships.

where everything is as previously defined except that the time at which the variables $(Y, x_1, \ldots, x_k)$ are taken is specified. Since the values change with t, the error term is also time-specific.

In order to provide adequate estimates of Y, conditions on the errors e_t must be imposed in such a model. First, the errors must average to zero. This is intuitive: if the errors have a non-zero mean it implies that the model is not capturing a global, intertemporal effect. The second condition is that errors should be uncorrelated with each other. A violation of this principle could mean that some seasonal or cyclical effect is being ignored by the model. Third, the errors should not be correlated with the predictors $x_{i,t}$. Indeed, if the residuals systematically vary when a specific independent variable increases, then the model is not appropriately capturing its effect. Under these assumptions the multiple regression model will provide adequate forecasts. Note that the additional conditions of normally distributed errors and constant error variance are necessary to the production of prediction intervals.[5]

Once the selection of explanatory variables has been made, the model must be optimized. The parameters $\beta_1, \beta_2, \ldots, \beta_k$ are chosen jointly to minimize the sum of squared errors (SSE). Since the SSE is a measure of fit, it is important to then evaluate the model's *predictive* capabilities when trying to forecast the values of $Y_{m+1}, Y_{m+2}, \ldots, Y_{m+h}$ with data up to period m. Also note that the model produces a prediction of Y_{m+j} as a linear function of other variables *taken at time* $m+j$. If for example x_1 is gross domestic product (GDP), then the GDP figure at time $m+j$ is required to generate the estimation. The predictors need to be independently forecasted beforehand so they can be used as the regression's input. Typically, independent variables include macroeconomic or demographic quantities such as GDP, population and interest rates.

Despite the model being linear in form, it is possible to introduce non-linear effects by transforming variables. One of the most common transformations consists in taking the natural logarithm of a variable so that its movements can be interpreted in percentages. For example, in the simple regression

$$Y_t = \beta_0 + \beta_1 x_{1,t} + e_t$$

β_1 can be interpreted as the quantity by which Y_t should increase if $x_{1,t}$ rises by 1 when all other factors are being held constant. If instead the following model is observed,

$$Y_t = \beta_0 + \beta_1 \ln(x_{1,t}) + e_t,$$

then β_1 is the amount by which Y_t increases when $x_{1,t}$ rises by 1%, all other things being equal. It is also possible to take the log of Y to express the change in percentage.[6]

5 These are not to be confused with confidence intervals. While similar in concept, confidence intervals are used when estimating the mean of a population and prediction intervals are to be used when predicting the value a random variable will take. For more on prediction intervals, see *Forecasting: principles and practice* by Rob J. Hyndman and George Athanasopoulos (2014).

6 For more on logarithm transformations and non-linear effects in general, see Chapter 2 of *Introductory Econometrics: A Modern Approach* by Jeffrey M. Wooldridge.

Alternatively, one can take other functions of the variables such as $x_{1,t}^2$ or $\sqrt{x_{1,t}}$, but it should be noted that a model's complexity does not guaranty its quality.

It is common in the aviation industry to use fully or partially logarithmized linear equations to produce forecasts. The International Civil Aviation Organization (ICAO) uses a full logarithmic model to forecast revenue passenger kilometers (RPKs) using GDP and yields.[7] The model is of the form:

$$\ln RPK_t = \beta_0 + \beta_1 \ln GDP_t + \beta_2 \ln yield_t + e_t$$

where β_1 and β_2, the coefficients of the regression, can be interpreted as the variation (in %) that occurs on RPK when the associated variable increases by one percent. Hence a rise of one percent in yield will engender a β_2% change in revenue passenger kilometers, all other things being equal.

Table 4: Typical long-term passenger forecasting regression models

	Using GDP (x_1)	Using GDP (x_1) and population (x_2)
Linear	$Y_t = \beta_0 + \beta_1 x_{1,t} + e_t$	$Y_t = \beta_0 + \beta_1 x_{1,t} + \beta_2 x_{2,t} + e_t$
Log-linear	$\ln Y_t = \beta_0 + \beta_1 x_{1,t} + e_t$	$\ln Y_t = \beta_0 + \beta_1 x_{1,t} + \beta_2 x_{2,t} + e_t$
Linear-log	$Y_t = \beta_0 + \beta_1 \ln x_{1,t} + e_t$	$Y_t = \beta_0 + \beta_1 \ln x_{1,t} + \beta_2 \ln x_{2,t} + e_t$
Log-log	$\ln Y_t = \beta_0 + \beta_1 \ln x_{1,t} + e_t$	$\ln Y_t = \beta_0 + \beta_1 \ln x_{1,t} + \beta_2 \ln x_{2,t} + e_t$

These standard regressions can be used to model passenger volumes using combinations of gross domestic product and population as explanatory variables. Table 4 illustrates the set of models used to produce medium- to long-term forecasts. The final figures are a function of GDP or GDP and population, and the relationship between these variables and the passenger figures depend on the form of the model.

Kenza model

It is possible to use economic data to formulate predictions without resorting to classic regressions. Typically, the variables in a linear regression are chosen from the intuition brought about by theoretical models. The procedure described here involves keeping the theoretical model's equations, adjusting them to the specificities of the data and using this calibrated model to generate forecasts. In essence, the result can be described as a finely tuned simulation extrapolated h years forward.

The Kenza model is specifically designed to forecast air traffic. Let the group of individuals above a certain income threshold be designated as *potential passengers*

7 ICAO Document 8991 AT/722/3

based on a population distribution. Let also the group of individuals from these potential passengers who actually travelled be called *actual passengers*. Given this, the model can be written as:

$$D_t = k_1 \cdot P_t \cdot F(k_2 \cdot \rho_t)$$

where D_t is the quantity of passengers[8] at time t, P_t is the population for period t and ρ_t is the average normalized price level for year t. ρ_t is defined as follows:

$$\rho_t = T_t \times \frac{GDP_{t_0}/P_{t_0}}{GDP_t/P_t}$$

where T_t is the indexed fare at time t and t_0 denotes the reference year. The function F plots the percentage of total population with income superior to the income threshold designated by $k_2 \cdot \rho_t$. It is referred to as the Kenza distribution.[9]

Note that the model parameters k_1 and k_2 are not values of a variable for years 1 and 2. These constants are assumed to be invariant through time, and are chosen to minimize

$$\sum_{t=1}^{m} e_t^2 = \sum_{t=1}^{m} [D_t - k_1 \cdot P_t \cdot F(k_2 \cdot \rho_t)]^2,$$

the sum of squared errors for the documented years ($t = 1, 2, \ldots, m$). While k_1 can be interpreted as the ratio of actual to potential passengers, k_2 is a multiplicative constant to the adjusted fare ρ_t. For any real value of k_2 the optimal value of k_1 is

$$k_1^* = \frac{\sum_{t=1}^{m} D_t \cdot P_t \cdot F(k_2 \cdot \rho_t)}{\sum_{t=1}^{m} (P_t \cdot F(k_2 \cdot \rho_t))^2},$$

which can be obtained by taking the first order condition of the minimization problem. See Annex 1 for a demonstration of this result. The second constant's optimal value can be found with the use of the bisection algorithm, a procedure that iteratively compares halves of the remaining domain in search for an optimal value. Although it can be slow compared to other computer-assisted optimization processes, it is a very common and robust optimization algorithm.

[8] A passenger is counted once per arrival and once per departure with the exception of direct transits. In this specific case, a passenger will arrive at an airport and depart from it being counted only once. *Passengers* are not to be confused with *actual consumers*.

[9] See Annex 2 for more information on Kenza distribution and elasticities.

Model selection

Once the data has been processed and several predictions have been generated using the aforementioned methods, how does one choose the best estimates? Forecasting is an exercise in uncertainty, meaning that there is no perfect *ex ante* measure of a model's predictive capabilities. Nevertheless, there are tools that help determine the appropriate model.

The easiest method is to observe the models' properties and remove any that are inadequate. For example, the errors can provide valuable insight. The obtained residuals should be somewhat random-looking with a zero mean. If a trend or a pattern subsists in the residuals, then there is some predictable effect that the model has not captured, and the prediction intervals will not be reliable. Typically, this kind of elimination procedure is insufficient to narrow the list down to a single model. Additionally, it is usually redundant when using more advanced methods.

A traditional way to choose between models is to compare fit measures. The model chosen to compute the forecast is the one that best fits the available data. While this is easy and intuitive, it only makes sense under strong assumptions. First, one needs to assume that there is an underlying model which dictates the movements of the variable of study and that this model will remain unchanged throughout the forecast horizon. This means that the information contained in the available data can be used to reliably model the variable's movements. The second required assumption is one of completeness. It says that *all* the information pertaining to the underlying model can be found in the available data. If this is not the case, fit measures and forecast accuracy remain unrelated and one may end up with a model similar to high-degree polynomials: amazing fit performance coupled with poor forecast accuracy.

Fortunately, another approach exists which allows us to drop this rather strong second assumption. Cross-validation involves testing the forecasting capabilities of the model. It means using sections of the dataset to produce forecasts for other sections. The differences between the observed data and the estimates produced by the model are referred to as "predicted residuals," and these provide a good measure of the reliability of a model's forecasts. Where fit measures show *how well the model sticks to the data being used in its construction*, cross-validation measures give an indication as to *how well the model sticks to the data that it is trying to predict*. Leave-k-out or k-fold cross-validation measures can be computed, with k depending on the length of the available data series.[10] There also exist widely used statistics which are constructed to help gauge a model's accuracy. Examples include Schwarz's Bayesian Information Criterion (BIC), AIC and its corrected version for small samples (AICc).[11]

[10] Many more cross-validation tests exist. See Arlot and Celisse (2010) for a rather thorough review of these measures.

[11] Idem.

Table 5: Summary of ACI forecasting models

	Models
Short term	Decomposition models Exponential smoothing ARIMA
Medium and long term	Linear regression Log-linear, Linear-log, Log-log Trend projection Kenza model

It should be noted that the predicted residuals represent the errors that can be reasonably expected from the forecast given the first assumption (i.e., that the future can be predicted reliably from the past and present observations). While it is impossible to consistently choose the model that best fits the yet unobserved future values, it is important to pick the model which best performs in terms of forecast accuracy.

References

International Civil Aviation Organization (ICAO). "Manual on Air Traffic Forecasting." 3 (2006).

Arlot, Sylvain, and Alain Celisse. "A survey of cross-validation procedures for model selection." *Statistics surveys* 4 (2010): 40-79.

Akaike, Hirotugu. "An information criterion (AIC)." *Math Sci* 14.153 (1976): 5-9.

Schwarz, Gideon. "Estimating the dimension of a model." *The annals of statistics* 6.2 (1978): 461-464.

Box, G. E. P., and G. M. Jenkins. "Reinsel GC." *Time series analysis forecasting and control. Wiley* (2008).

Gardner, Everette S. "Exponential smoothing: The state of the art." *Journal of forecasting* 4.1 (1985): 1-28.

Cleveland, Robert B., et al. "STL: A seasonal-trend decomposition procedure based on loess." *Journal of Official Statistics* 6.1 (1990): 3-73.

Theodosiou, Marina. "Forecasting monthly and quarterly time series using STL decomposition." *International Journal of Forecasting* 27.4 (2011): 1178-1195.

Sallier, Daniel. "The 2nd Kenza Law of Demand: A New Qualitative and Quantitative Approach for Assessing Long-term Developments of the Demand for Leisure Destinations." *European Transport Conference, 2010*. (2010).

Mahmoud, Essam, and C. Carl Pegels. "An approach for selecting times series forecasting models." *International Journal of Operations & Production Management* 10.3 (1990): 50-60.

Wooldridge, Jeffrey. *Introductory econometrics: A modern approach*. Nelson Education, (2015).

Hyndman, Rob J., and George Athanasopoulos. *Forecasting: principles and practice*. OTexts, (2014).

Annex 1 – Deriving the optimal value of k_1

The value of k_1 is obtained by minimizing the sum of squared errors.

$$\sum_{t=1}^{m} e_t^2 = \sum_{t=1}^{m} [D_t - k_1 \cdot P_t \cdot F(k_2 \cdot \rho_t)]^2$$

By developing the square we obtain

$$\sum_{t=1}^{m} e_t^2 = \sum_{t=1}^{m} [D_t{}^2] - 2\sum_{t=1}^{m} [D_t \cdot k_1 \cdot P_t \cdot F(k_2 \cdot \rho_t)] + \sum_{t=1}^{m} [k_1 \cdot P_t \cdot F(k_2 \cdot \rho_t)]^2.$$

We then optimize for k_1 and write the first order condition

$$FOC_{k_1}: \frac{\partial \sum_{t=1}^{m} [D_t{}^2] - 2k_1 \cdot \sum_{t=1}^{m} [D_t \cdot P_t \cdot F(k_2 \cdot \rho_t)] + \sum_{t=1}^{m} [k_1 \cdot P_t \cdot F(k_2 \cdot \rho_t)]^2}{\partial k_1} = 0,$$

which equates to

$$-2\sum_{t=1}^{m} [D_t \cdot P_t \cdot F(k_2 \cdot \rho_t)] + 2k_1 \cdot \sum_{t=1}^{m} [P_t \cdot F(k_2 \cdot \rho_t)] \cdot [P_t \cdot F(k_2 \cdot \rho_t)] = 0.$$

After simple manipulations, we obtain

$$k_1^* = \frac{\sum_{t=1}^{m} [D_t \cdot P_t \cdot F(k_2 \cdot \rho_t)]}{\sum_{t=1}^{m} [P_t \cdot F(k_2 \cdot \rho_t)]^2},$$

the optimal value of k_1.

Annex 2 – Kenza distribution and elasticities

The intrinsic elasticity of a Kenza distribution is defined as:

$$\varepsilon_K(r) = \frac{dF(r)}{F(r)} \cdot \frac{r}{dr} \leq 0$$

with r being the normalization quantity by which prices are divided to obtain ρ_t. It can be proven that the intrinsic elasticity is completely independent of the normalization quantity being used (GDP per capita, average individual income, median individual income, etc.). Assume that the demand is modelled by the first Kenza law of demand,

$$D_t = P_t \cdot k_1 \cdot F(k_2 \cdot \rho_t),$$

then the demand elasticity to price p_t is equal to

$$\varepsilon_{p_t} = \varepsilon_K(k_2 \cdot \rho).$$

The demand elasticity to the normalization quantity r (e.g., GDP per capita or average revenue per capita) is equal to

$$\varepsilon_r = -\varepsilon_K(k_2 \cdot \rho),$$

so demand elasticities directly derive from the Kenza intrinsic elasticity.

Analysis of actual Kenza distributions for several countries shows large sections featuring an almost linear relationship between the intrinsic elasticity and the corresponding potential passengers (sorted by decreasing levels of income). Thus, the 1st Kenza law of demand "naturally" takes into account the market's long term, rather linear maturation process (demand elasticity to price decreases in absolute value). It should be pointed out that the market maturation process is not related to the market being studied. It means that it exclusively results from the way individual incomes are distributed in the considered population. The only market-related phenomenon is how fast the potential passenger segment grows or decreases with time. One of the main characteristics of the model is based on the Kenza distribution, which illustrates the (nonlinear) link between the individual revenue and the demand curve. The simplified Kenza function, based follows a linearization of the Kenza distribution of income, can be defined as

$$D_t = P_t \cdot k_1 \cdot [a(k_2 \cdot \rho_t) + b],$$

or more simply as $\frac{D_t}{P_t} = A \cdot \rho_t + B$.

This could be considered as a very simple econometric linear model. The difference between the full and simplified Kenza models provides intuition as to the added value of taking into account the Kenza distribution function.

AIRPORTS COUNCIL INTERNATIONAL

ACI World
PO Box 302
800 Rue du Square Victoria
Montreal, Quebec
H4Z 1G8 Canada

www.aci.aero
aci@aci.aero
Tel. +1 514 373 1200
Fax. +1 514 373 1201