NEW DIRECTIONS FOR FORECASTING AIR TRAVEL PASSENGER DEMAND

Donald S. Garvett
Nawal K. Taneja

Simpson

MIT

DEPARTMENT OF AERONAUTICS & ASTRONAUTICS

FLIGHT TRANSPORTATION LABORATORY
Cambridge, Mass. 02139

R74-3

July 1974

FLIGHT TRANSPORTATION LABORATORY
DEPARTMENT OF AERONAUTICS AND ASTRONAUTICS
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

FTL R-74-3

July 1974

NEW DIRECTIONS FOR FORECASTING AIR TRAVEL PASSENGER DEMAND

Donald S. Garvett
Nawal K. Taneja

Part of this work was performed under a Grant Number GI-39695 from the National Science Foundation, Washington, D.C.

Table of Contents

Table of Symbols and Notation

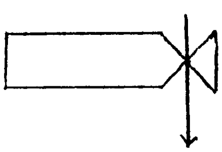
OLS	Ordinary Least Squares
GLS	Generalized Least Squares
COV(X,Y)	Covariance of X and Y
VAR(X)	Variance of X
— (Bar)	Average or mean, e.g. Average Value of X = $\bar{X}$
' (Prime)	Extrapolated or updated value of a quantity, e.g. updated value of X = X'

bold faced capitals represent matrices, e.g. $\mathbf{E}$, $\mathbf{\Phi}$

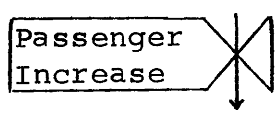
bold faced lower case symbols represent column vectors, e.g. $\mathbf{x}$, $\mathbf{y}$

∂ partial derivative, e.g. the partial derivative of x with respect to y = $\frac{\partial x}{\partial y}$

^ (hat) Estimated value of a quantity, e.g. estimated value of X = $\hat{X}$



Symbol for a rate in feedback model flow diagram, e.g. the rate at which passenger increase can be represented as:



T Transpose operator:

vector transpose: $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix}$ $\mathbf{x}^T = [x_1, x_2, \ldots, x_N]$

matrix transpose: $\mathbf{E} = \begin{bmatrix} E_{11} & E_{12} \\ E_{21} & E_{22} \end{bmatrix}$ $\mathbf{E}^T = \begin{bmatrix} E_{11} & E_{21} \\ E_{12} & E_{22} \end{bmatrix}$

matrix-vector operation: e.g.

$$\mathbf{Eb} = \begin{bmatrix} E_{11} & E_{12} \\ E_{21} & E_{22} \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} E_{11}b_1 + E_{12}b_2 \\ E_{21}b_1 + E_{22}b_2 \end{bmatrix}$$

vector-vector operations:

$$\mathbf{a}^T\mathbf{b} = \begin{bmatrix} a_1 & a_2 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} a_1b_1 + a_2b_2 \end{bmatrix}$$

$$\mathbf{ab}^T = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} \begin{bmatrix} b_1 & b_2 \end{bmatrix} = \begin{bmatrix} a_1b_1 & a_1b_2 \\ a_2b_1 & a_2b_2 \end{bmatrix}$$

$\sum$	Summation, e.g. $\sum_{i=1}^{N} X_i = X_1 + X_2 + \ldots X_N$
Π	Product, e.g. $\prod_{i=1}^{3} X_i = X_1X_2X_3$
$\approx$	Approximately equals
GNP	Gross National Product
E_X^Y	Point Elasticity of Y with respect to X $= \frac{\partial Y}{\partial X} \frac{X}{Y}$
ln	Natural logarithm
e	Base of natural logaritm ≈ 2.7182818

LOS	Level of Service
ϵ	Element of
δij	Kronecker delta: $\delta ij = \begin{cases} 1 & i=j \\ 0 & i \neq j \end{cases}$ this results in a matrix of the following form: $\begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & 0 & \cdots & 1 \end{bmatrix}$
$\forall$	All elements of
P(X)	Probability of event X occurring
P(X|Y)	Conditional probability of event X occuring given that event Y has occurred
P(X,Y)	Joint probability of events X and Y occurring
ARIMA	Auto-Regressive Integrated Moving Average
π	Circumference of a circle divided by its diameter 3.14159
(±X)	Lagged quantity - the quantity Y lagged 4 time periods = Y(-4)
SER	Standard Error of Regression
CRSQ	Corrected R-Squared Statistic

σ	Standard deviation
$\sigma 2$	Variance
DW	Durbin-Watson Statistic
t	Student's t statistic
μ	Mean or average value
Randu	Random number generator in Fortran (yields uniformly distributed variable between 0 and 1)
Gauss	Random number generator in Fortran (yields Gaussian distributed variable)
TUBSIM	Trasportation user behavioral simulation model
$\nabla^d Z_t$	The d^{th} difference of the raw time series, Z_t

e.g. $\nabla Z_t = Z_t - Z_{t-1}$

$\nabla^2 Z_t = Z_t - Z_{t-1}$

etc.

List of Figures

List of Figures (Continued)

List of Tables

Chapter 1

Introduction

While few will disagree that sound forecasts are an essential prerequisite to rational transportation planning and analysis, the making of these forecasts has become a complex problem with the broadening of the scope and variety of transportation decisions. Until recently, the forecasting methods available addressed the issues which were important a couple of decades ago. These methods attempted to predict the amount and in some cases character of travel to be used in designing major highways, transit facilities, seaport facilities, and airports. However, today's issues to be addressed in transportation are much broader and more complex. For example, in the modern process of transportation planning, the decision-maker is concerned with the broad range of social, economic and environmental effects, equity issues, wider range of options including not building major facilities, resource constraints such as energy, and increased public participation in the planning process in general.

The complexity of the problem has necessitated the planner's developing improved methods of forecasting the demand for transportation at all levels and by all modes. While significant contributions have been made recently to the development of improved methods in forecasting, we are still a long way from possessing tools which provide our decision-makers with more effective, that is, more useful, accurate and timely information.

The purpose of this report is to present a very brief overview of the current and emerging air transporation forecasting methods with the aim of identifying areas which need further research. Throughout the report, the object is to indicate future directions for research into transportation forecasting methods which are more responsive to today's issues. For example, it is clear from reviewing the literature that tremendous improvements in travel forecasting methods can be achieved through deeper understanding of the traveller's behavior, under a range of conditions, development of models which are more policy-responsive and development of improved data bases.

Peculiarities of the airline industry and aviation in general cause many standard techniques of economic and managerial analyses to break down. Air travel demand is unique in that even the sophisticated techniques developed by urban transportation analysts are often not directly applicable to modelling the demand for air transportation. Econometricians usually do not have specific training in air transportation. Airline managers, on the other hand, quite often do not have the technical beckground necessary to fully understand many highly detailed and complex models. In order to develop sophisticated yet user-oriented models, an analyst must have background in several areas. It is hoped that the material presented in this report will help bridge the gap between managerial and technical personnel and provide some new directions for air travel demand modelling.

Generally speaking, there are two broad categories of forecasting methods. The quantitative group is composed of techniques which rely on the existence of historical data, and which assume that the historical trend will be expected to continue in the future. This group is further divided into two classes, time-series methods and causal methods. The quantitative techniques are by far the most widely used and contain such popular methodologies as moving averages, classical decomposition analysis, spectral analysis, adaptive filtering and Box-Jenkins methods under the category of time-series analysis. The causal methods contain such favorites as modelling classical consumer behavior through regression models and more recent applications in transportation demand analysis of bayesian analysis, markov chains, input-output analysis, simulation methods and control theory models.

The second group of forecasting methods is qualitative in nature. The techniques in this group are used when none or very little historical data exists, or when the underlying trend of the historical data is expected to change. Qualitative techniques have in general been applied to project future technological developments and their impacts are described in literature as "technological forecasting methods." The group is further divided into two classes, exploratory and normative methods. The exploratory methods start with today's knowledge and its orientation and trends and seek to predict what will happen in the future and when. On the other hand, normative methods seek first to assess the organization's goals and objectives and then work backwards to identify new developments which will most likely lead to the achievement of these goals. Familiar examples of exploratory methods are the envelope, logistic or S-curve, the Delphi technique and morphological analysis. Examples of methods used to perform normative forecasting are relevance trees and cross-impact analysis.

Although this classification scheme is consistent with the way that many forecasters might differentiate models, it is by no means unique. Other and perhaps better classification schemes exist. For the purposes of this report we will not attempt to define a particular classification but present five broad areas which show the greatest potential for improving our capabilities of modelling the demand for air transportation. These areas are: technological forecasting, time-series models, control theory models, econometric models and simulation models. Each of the general techniques are reviewed, and specific examples are presented where relevant. Excessive mathematical detail was avoided in order to make this work easily understandable by managers and others who might not have a rigorous analytical background. Since a number of models discussed in the report require extensive computer modelling, we have included a few computer programs in the appendices to make the report more user-oriented.

Chapter II

Technological Forecasting

As defined in the introduction the quantitative techniques rely on the existence of appropriate and sufficient historical data upon which to derive a forecast and assume that some underlying trend or trends in the data will be continued into the future. The qualitative techniques, on the other hand, are used when no historical data exists or when the underlying trend of the historical data is expected to change. Qualitative forecasting is most often concerned with what events are likely to occur in the future rather than the specific time of the event. Because qualitative techniques have so far been most widely used to project future technological configurations and their impact, they are generally described in the literature as "technological forecasting methods." There are at least two dozen principle technological forecasting methods and perhaps as many as one hundred variations around the basic methods. In this chapter we will attempt to describe the more well known of these techniques.

In technological forecasting, one is generally interested in the prediction or determination of the feasible or desirable characteristics of performance parameters in future technologies. The fundamental questions posed in technological forecasting are: What is possible? What is expected? What is desired or intended?

In general, technological forecasting techniques can be divided into two broad categories: exploratory and normative. Exploratory techniques attempt to generate new information about future systems and performance or to simulate the outcomes of anticipated events. They are used to broaden the analyst's knowledge of what can be expected or what might possibly happen. Exploratory techniques which attempt to generate new information can further be divided into two types: extrapolative and speculative. An example of an extrapolative technique is the envelope, logistic or S-curve. Speculative techniques are more often used to address the question, "what is possible?" and rely to a greater extent on intuition. The Delphi Technique and Morphological Analysis can be included in this category.

Within the scope of exploratory methods, the analyst can simulate outcomes assuming different combinations of events. Examples of appropriate simulation models are: input-output analysis, scenario writing, and cross-impact analysis.

With the normative models, the purpose is the same as with exploratory models; the generation of new information or the simulation of outcomes. However, these activities take place within the context of achieving certain desires or structural relationships. The aim is to identify those critical linkages or steps which must be made or taken in order to reach the desired end state. In addition to the methods already mentioned under exploratory techniques, other methods such as relevance trees, decision theory and dynamic modeling can be used to perform normative forecasting.

The simplest method of obtaining a technological forecast is to get the opinion of an recognized expert. However, because of the inevitable complexities of technological problems, the opinion of a single expert is of limited value. The next step is to obtain the advice and concensus if possible of a panel of experts. In the case of transportation such a panel should be interdisciplinary. The group consisting of engineers, economists, demographers, transportation experts, city planners and political scientists would bring their respective areas of expertise to bear on the transportation problem at hand. For example, what impact would the development of domestic air transportation have on population dispersion?

A weakness of the panel approach is that it can be biased by the persuasiveness of individual members who may or may not have a valid argument. To overcome such psychological interferences which tend to reduce the value of forecasts reached by the group, Olaf Helmer at the RAND Corporation developed the Delphi Technique. The aim of the Delphi Technique is to develop a carefully designed program of sequential individual interrogations, usually conducted by questionnaries, interspersed with information and judgment feedback derived from the consensus of the earlier parts. The idea is that, through successive trials, the spread of forecasts is reduced.

In a survey on air transportation developments reported by McDonnell Douglas (1970) a questionnaire was sent to 304 experts representing the decision-making levels of management from four major segments in the air transportation industry.[1] Using the Delphi Technique each participant was asked to make an anonymous forecast of future air transportation developments and the date each event in question would happen. The participants were given a composite feedback based on the analysis of the tabulated forecast. Based on this feedback, the same experts were then asked to make a second and final forecast. They could either take a firm stand on their original opinion or they could revise them. More than half of the participants revised their forecasts based on the feedback. The areas covered in the survey were advances in air transportation technology, development of air cargo, passenger preference for air transportation, seat capacities of future aircraft, and new markets for air transportation usage. Table 2.1 shows the summary of the results.

Table 2.1

RESULTS OF A SURVEY ON AIR TRANSPORTATION DEVELOPMENTS

Item	One fourth of the respondents thought by this date	One half thought by this date	Three fourths thought by this date
1. A commercial aircraft powered by a nuclear propulsion system will be operational.	(1990)	(1995)	(2000)
2. Revenue from commercial air cargo on certified airlines will equal passenger revenue.	(1985)	(1990)	(2000)
3. A large, separate cargo airport for a major population center will be operational.	(1980)	(1982)	(1990)
4a. Free world certified and chartered revenue passenger miles will reach one trillion.	(1990)	(1990)	(2000)
4b. Free world certified and chartered revenue passenger miles will reach two trillion.	(2000)	(2000+)	(2000+)
5. A 1000 passenger aircraft will be introduced into commercial service.	(1985)	(1990)	(1995)
6a. A cargo aircraft having a 500 ton cargo payload capability will be introduced into service.	(1985)	(1990)	(1995)
6b. A cargo aircraft having a 1500 ton cargo payload capability will be introduced into service.	(2000)	(2000+)	(2000+)
7. Twenty-five percent of the commuters who travel one-way distances greater than 15 miles will use air transportation.	(2000)	(2000+)	
8. Twenty-five percent or more of the passengers departing from a major air terminal will arrive at the terminal by a feeder airline.	(1980)	(1985)	(1990)
9. You and your associates will normally use air transportation to travel to conferences with your business associates located in different parts of the same megalopolis.	(1990)	(2000+)	

Table 2.1 (Continued)

10.	New exotic materials such as boron filament and beryllium will be commercially competitive and in general usage to partially replace the conventional aluminum, titanium, and steel aircraft structure.	(1980)	(1985)	(1990)

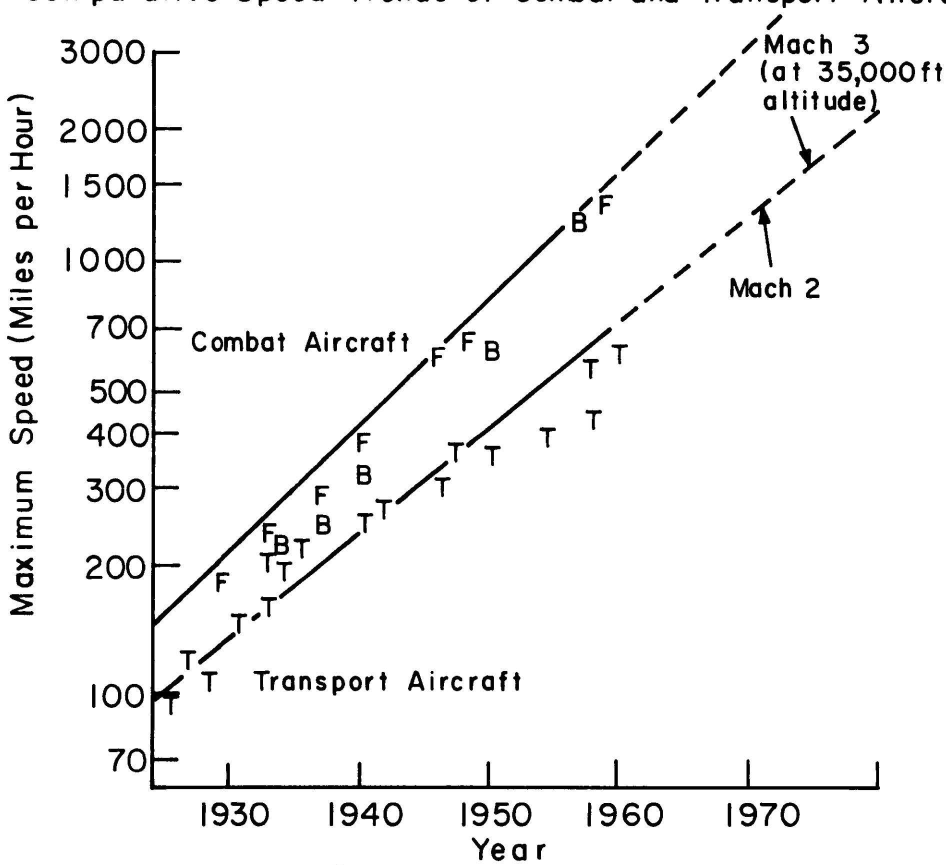
Source: McDonnell Douglas, "McDonnell Douglas Asked the Experts Their Opinion of Important Future Air Transportation Developments," September 1970.

Another intuitive approach called forecasting by analogy, is to compare some developing technology with some similar technology in the past. Joseph Martino suggests that analogous situations be analyzed for comparability along the following lines: technological, economic, managerial, political, social, cultural, intellectual, religious/ethical and ecological.[2] As an example of a formal historical analogy, Martino discusses technology transfer comparing the railroad industry and the more recent space industry along the nine dimensions given above. His forecast is as follows:

> During the construction and development of the railroads in America, between 1830 and 1870, considerable new technology was developed. With the termination of the railroad building era in 1870 or shortly thereafter, many of the engineers who had carried out the work...were no longer needed for this task. They found other employment solving non-railroad problems. In the solution of non-railroad problems, they tended to use the railroad technology they were familiar with.... During the course of the American space program also, much new technology was developed. By 1970 it was evident that the space program would proceed at a slower pace.... The engineers displaced from the space program and finding employment in other fields by analogy with the experience of railroads, will transfer a significant amount of space technology to these other fields.[3]

A useful variant of the forecasting by analogy approach discussed above is known as the precursive method. This method relies on the assumption that technical progress in one area lags another by a given length of time. The leading indicator is usually called the precursive indicator. For example, this method has been used in the past to forecast the maximum speed of commercial aircraft from its relationship with the maximum speed of military aircraft. In 1961 Lenz found that the speed of commerical aircraft followed the speed of military aircraft by 6 years in the 1920's and eleven years in the 1950's. See Figure 2.1.[4] The major difficulty in using this approach lies

Comparative Speed Trends of Combat and Transport Aircraft



Source: R.C. Lenz, "Technological Forecasting," ASD-TDR-62-414, Wright-Patterson AFB, Ohio: Air Force Systems Command, June 1962.

H. W. Lanford, "Technological Forecasting Methodologies" American Management Association, 1972. p. 94

Figure 2.1

in defining the time lag. The credibility of this approach depends very highly on the establishment of a logical connection between the two trends. In the case of forecasting the maximum speed of transport aircraft from military aircraft this connection is at least plausible.

Another example of forecasting by analogy is the application of Pearl's Law on biological growth applied to technological development. According to Pearl, the increase of population in a given area follows a pattern similar to the increase of biological cells. This may be expressed mathematically as:[5]

$$y = \frac{L}{1+ae^{-bx}}$$

where y = accumulated information (state of knowledge) at time; y_0=initial value; y_1=value after one time unit

L = upper limit of information (due to constraints)

x = time; at y_0, x=0

a = constant (dimensionless); $a = \frac{L}{y_0} - 1$

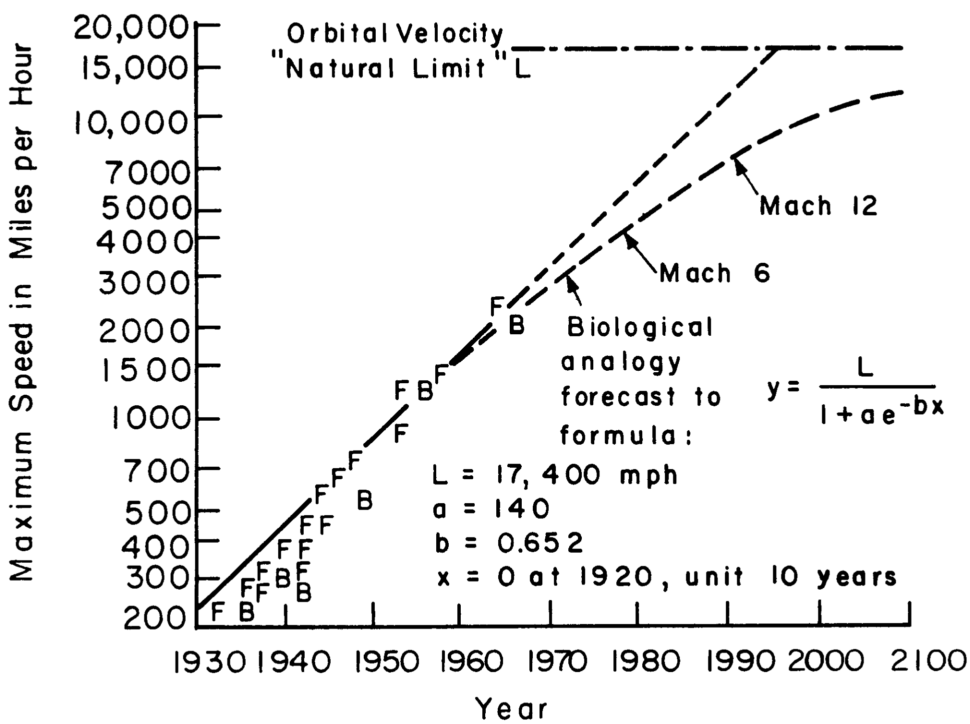
b = constant (per time unit); $b = -\log \frac{Ly_0 - y_0 y_1}{Ly_1 - y_0 y_1}$

e = basis of the natural logarithmic system (2.71828).

Lanford gives an example of this type of formulation applied to speed trends of U.S. aircraft shown in Figure 2.2. This provides a comparison of the exponential trend method with the biological analogy. Here Pearl's formula is applied with the extrapolation being asymptotic to the speed representing orbital velocity at 100 miles altitude.

Yet another widely used growth curve analogy is Gompertz's Law, often used to describe growth phenomena in such areas as income growth. The mathematical expression for Gompertz's Law is:[6]

Speed Trends of U.S. Aircraft



Source: R.C. Lenz, "Trend Extrapolation Techniques," a paper presented at the First Annual Technology and Management Conference, Industrial Management Center, Lake Placid, N. Y., May 1967.

H. W. Lanford, "Technological Forecasting Methodologies". American Management Association, 1972. p. 80

Figure 2. 2

$$y = Le^{-be^{-kx}}$$

where y = growth phenomena

L = limit (in the same units as the parameter y)

b,k = constants

x = time

e = 2.71828

This formula produces the familiar S-shaped growth curve in which the growth of increments of the logarithms decline at a constant rate.

There are numerous other growth curve techniques. The well known envelope curve technique, also called the S-curve or logistic curve calls for fitting an overall trend curve to a series of technological developments. The logistic curve can be fitted to a given product, technology or even some broader parameter. Robert Ayres shows an example of the application of this technique by fitting an envelope curve to the speed of transportation.[7] After plotting the curve of each individual technology, an envelope curve can be derived by connecting the tangents of of each as shown in Figure 2.3.[8] A common problem with this technique is that one expert's judgment on the behavior of the curve at the extremal point is as good as another. This problem can be somewhat mitigated by the limiting affect of practical or theoretical maximums such as speed of light and theoretical efficiency.

In addition to the intuitive technological forecasting techniques such as the Delphi and analytical techniques such as trend analysis and curve fitting, there is yet another set of techniques which relies on an orderly and systematic investigation of all opportunities at various levels for potential interactions among items in a forecasted set of occurences. This group consists of techniques such as scenario-writing, cross-impact

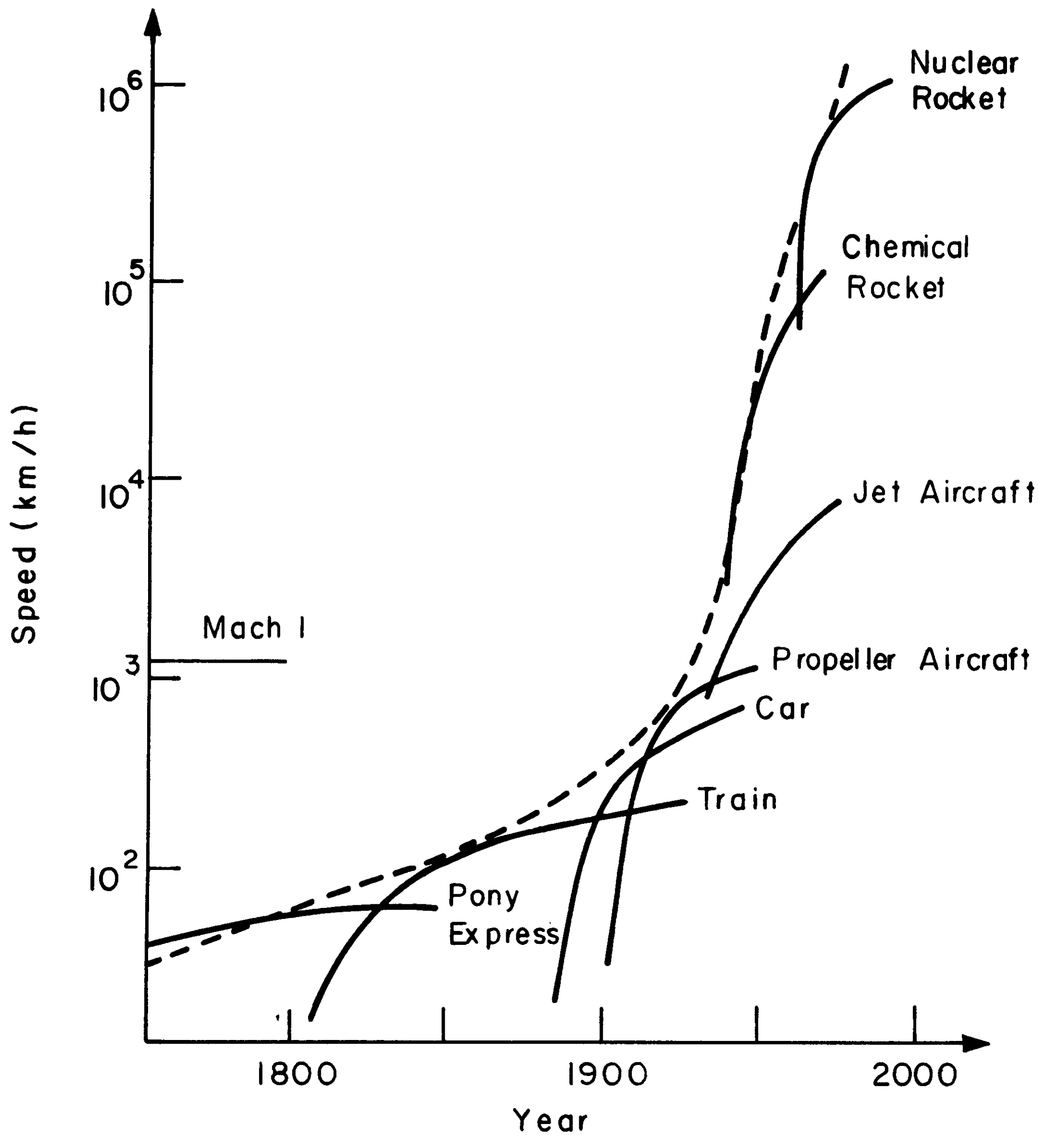


Figure 2.3 Envelope Curve Extrapolation

Source: Erich Jantsch, "Technological Planning and Social Futures," John Wiley & Sons, 1972. p. 85

analysis, morphological analysis, and relevance trees. This set of techniques has not yet been applied directly in the area of transportation forecasting. However, this is an area where further research would do much good.

Scenario-writing, a technique normally associated with the work of Anthony Wiener and Herman Kahn of the Hudson Institute is basically an attempt to view and combine various trends in some systematic manner. For small single steps the changes implied for the scenario are laid out and potential actions which have to be taken due to these changes are then systematically explored. The goal of scenario-writing is not necessarily to predict the future, but instead the purpose is to explore systematically the branching points dependent upon critical choices. Scenarios can be generated through the use of the Delphi technique and are useful in calling attention to a large number of possibilities which must be considered in the analysis of transportation systems involving the interaction of such forces as social, economic, and political.

An extension of the scenario approach which often uses the results of the Delphi technique is the Cross-Impact Analysis Method. This technique allows an orderly investigation of potential interactions among items in a forecasted set of occurrences. This method recognizes the causal links or chains between individual forecasts and allows the modification of scenarios to make them more internally consistent. Two versions have been developed. In one case, events are forecast together with estimates of timing and probability that the event will take place. Interactions are then considered. In the second version "point" forecasts (events) are divided into three broad categories: environment, applications and basic technologies. Further subdivision is possible. For each case causal networks are built of the events with new events being filled in where necessary. Additional networks, branching off from the original set can be added. Networks within the three categories are harmonized with respect to timing, estimates of probability, feasibility and desirability and brought into logical order.

Another orderly way of looking at things is the technique known as morphological analysis. Fritz Zwicky pioneered this technique in his exploratory studies in the field of jet engines. In essence, the technique attempts to break up the problem into its basic parameters, and then conceive of as many variations of each parameter as possible. Morphological analysis does not so much yield a forecast as it does open up possibilities.

The first step is to state the problem to be solved with the greatest possible precision. Next, all impinging parameters are identified. Third, the parameters are subdivided into distinguishable states. Lastly, rules for analyzing the combinations are formulated. One then proceeds through all possible combinations, eliminating those which are contradictions and seeking combinations which offer an opportunity for technological breakthrough. In a way similar to this, Fritz Zwicky was able to decompose a chemical jet characterized by 11 parameters into 36864 possible combinations of these eleven parameters. In the next step he reduced this number to 25344 after the systematic removal of internal inconsistencies. By selecting specific combinations, Zwicky was able to suggest a number of new conceptual inventions such as the "aeroduct," a ramjet utilizing the chemical energy of free radicals and excited molecules in the earth's upper atmosphere, and the "aeropulse" or "rocket pulse," which carries part of its own oxidizer and obtains the rest from the outside air during the negative pressure phase of its cyclic operation.[9]

Another taxonomical approach to identifying key linkages in the solution of technological problems or in forecasting where problems may be is that of the relevance tree. The method uses the ideas of decision theory to assess the desireability of future goals and to select those areas of technology whose development is necessary to the achievement of those goals. Objectives are broken down into successively small components each with its own branching possibilities. The outcome is that one is able to identify all objectives at any level and derive a quantitative value which would benefit from achievement for any particular branching route. A problem with relevance trees is that they require fairly precise ideas at all levels of activity. As a consequence they are generally used only for forecasting 10-15 years into the future.

The technological methods outlined in this chapter represent good potentials for new areas for the development of demand for air transportation. While some of these have been used, further research is needed before any of these can be implemented at the practical level. Finally, the reader is cautioned that the list of methods discussed in this chapter is by no means complete. Other methods such as the application of systems dynamics, game theory and decision analysis should also be considered.

Chapter III

Time-Series Analysis

The oldest and in some cases still the most widely used methods for forecasting the demand for transportation is time-series analysis, more simply known as trend-extrapolation. The method often used where time and data are limited, produces the forecast of a single variable, passengers carried or transported through the use of historical data for the particular variable. The historical data can be manipulated through the use of sophisticated smoothing techniques. Since time is used to reflect the impact of many variables, the method is only useful as long as there is no change in this basic trend.

Trend extrapolation is often thought of as a simple and a rough method of producing a forecast. However, before an analyst discards the use of trend analysis, he must keep two things in mind. First, while other methods may appear more appealing based on theoretical grounds, data may not be available to justify their use. For instance, how does one quantify social status as an explanatory parameter for overseas pleasure travel, or privacy as a factor in the choice of car as a mode of travel. Second, a model's simplicity is in the mind of its user. For example, recent applications of adaptive filtering methods, Box-Jenkins methods and spectral analysis can hardly be classified as "simple and rough" and their use has considerably increased the validity of trend-extrapolation.

Time-series analysis is especially useful in producing short-term forecasts. In particular, forecasts of monthly, weekly, daily and hourly variations can most easily be produced by using time-series models. These variations are required, for instance, to measure the impact and magnitude of peak loads. While in the past, the most common means for handling fluctuating patterns have been simple and exponential smoothing techniques, significant improvements have been achieved through the development of such techniques as adaptive filtering, Box-Jenkins methods and spectral analysis.

An Adaptive Forecasting Technique

Forecasting models are just representations of the world as viewed at a given period in time. As the world is constantly changing, it is only reasonable that models should also change. Models that change through time are denoted as adaptive techniques. They have the potential for yielding better forecasts; but associated with this improvement is a higher cost of estimation and development as well as usually increasing complexity.

In one type of adaptive forecasting method, the structure of a model can change through time. One might decide that a new variable has become important, wage-price controls should be accounted for explicitly, or that numerous functional changes should be made. In a model of automobile demand, dummy variables for Nixon's fiscal policies, automobile strikes (two different ones), and stell contract renegotiation were included. Other models have included even more out-of-place dummy variables. The addition of these very special purpose dummy variables (as opposed to more appropriate dummy variables such as seasonality) to a model indicates a fundamental weakness in it. It is rare that the addition of new variables will correct misspecification errors in a weak model.

The record has not been good for models that constantly change their structural form in adapting to new conditions. They generally do not do better than alternative models that have an unchanging structure. As a result, not very much work has been done on models with changing structures.

The second type of adaptive forecasting model is one in which the structure remains unchanging, but the values of the coefficients vary. Kalman filtering (which is presented in Chapter 4) is an adaptive technique. The gains (smoothing constants) are time-varying based on the variances and covariances (i.e. degree of uncertainty) associated with the state.

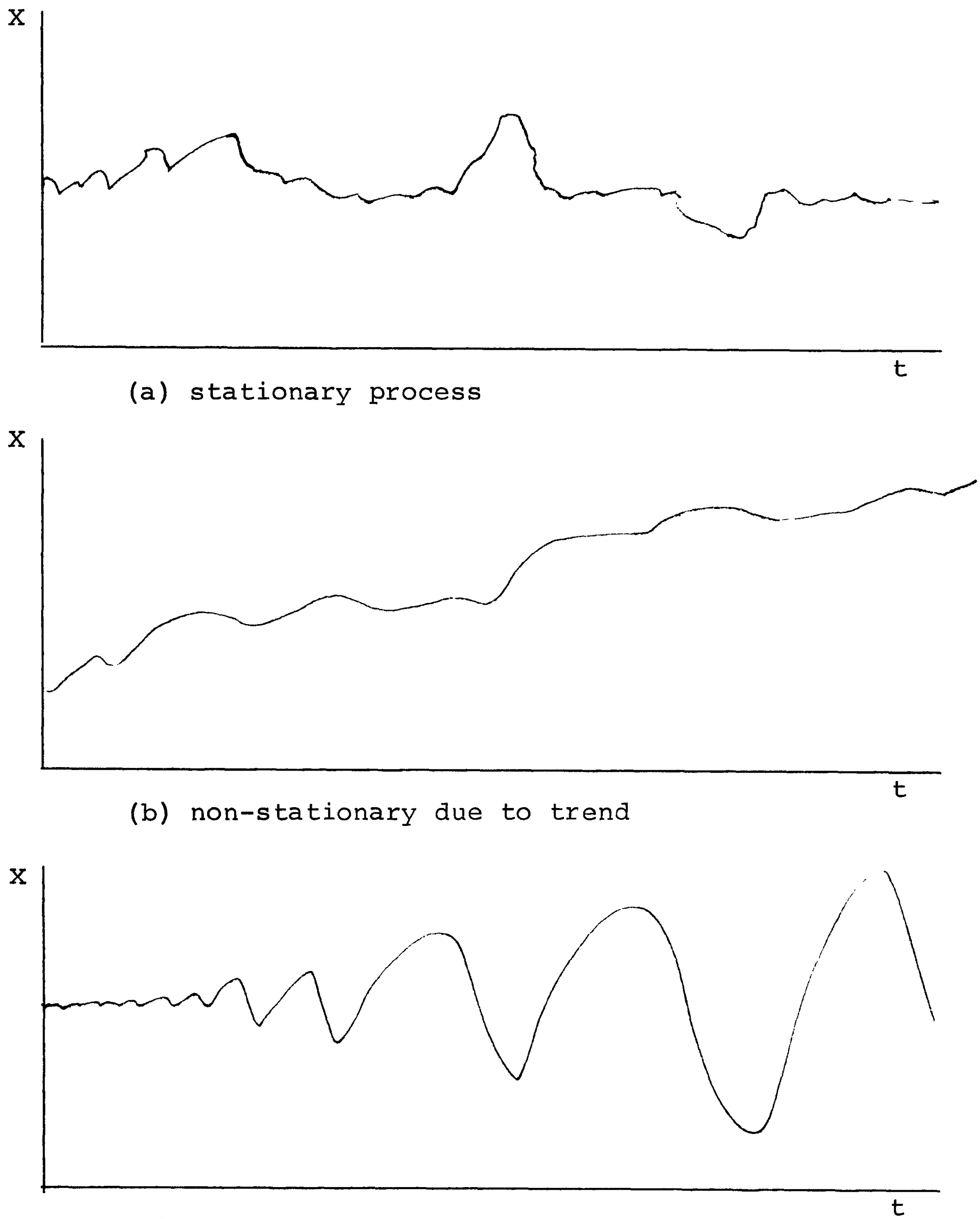
One often wishes to perform exponential smoothing on a series due to its simplicity. Better results can be achieved by using time-varying (adaptive) smoothing constants. Following are some examples of adaptive techniques as applied to exponential smoothing. These examples assume that the series to be evaluated is stationary. A stationary series is one whose mean and variance have no trend over time. Therefore a stationary series cannot have a time trend or other changes in average value. The constant variance assumption is necessary to insure responsiveness and stability. Most processes with unchanging means also have unchanging variances. Examples of stationary and non-stationary processes are shown in Figure 3.1.

One can often make a non-stationary series stationary by creating a series of the differences (e.g. create w(I) = X(I+1) - X(I) for all I) to remove trend in the mean; or to create a series of the differences of the logs (e.g. w(I) = log(X(I+1)) - log(X(I)) for all I) to remove trend in the mean and the variance. Alternatively, one could run a straight-line equation on the moving average of the series to eliminate the trend. If the series is not stationary, one can transform it until it becomes stationary or can estimate trend and seasonal terms with methods that are similar to those following:

Assume the following simple model:

$$X_{t+1} = \alpha_t X_t + (1-\alpha_t)\hat{X}_t \qquad (3.1)$$

X_t = observed value of series X at time t



(a) stationary process

(b) non-stationary due to trend

(c) non-stationary due to increasing variance

Figure 3.1. Stationary and non-stationary series

$\hat{X}_t$ = estimated value of X_t

α_t = time-varying smoothing constant

Given this model, one can recursively simulate the series if he has an initial value of $\hat{X}_t$. A suitable value of $\hat{X}_{t_{initial}}$ is the mean of the entire series of X_t.

One can simulate this series with different values of the smoothing constant (α). For the available data, let α_t be a constant (i.e. $\alpha_1 = \alpha_2 = \ldots = \alpha_N$). Choose the α that results in the minimum mean squared error. In a periodic, unconstrained adaptive technique, one will choose an initial α and repeat this process each time that a new observation is recorded. The smoothing constant is then re-evaluated each time that new information becomes available.

A periodic, constrained adaptive technique will choose an initial value of α in the same way as a periodic, unconstrained model will. However, a new value of α will not be calculated when new observations are made unless certain conditions are met. A typical condition might be to adjust α if the error of the forecast of the latest observation exceeds three times the standard deviation of the series. The adjustment might take many forms. A suitable adjustment could be to adjust α five percent in the direction that reduces the error. This technique will not give as good results as the unconstrained method, but it requires considerably less computer time to perform.

The final and most sophisticated adaptive smoothing technique that will be presented is a continuous, unconstrained method. It has the greatest computational requirements of the three methods that are discussed in this section, but it is also potentially the most accurate. It updates the smoothing constant (α) by using the forecast error (E_t) to determine how well the series tracks itself.

$$E_t = X_t - \hat{X}_t \tag{3.2}$$

$\bar{E}_t$ = smoothed forecast error

$$\bar{E}_t = (1-\gamma)\bar{E}_{t-1} + \gamma E_t \tag{3.3}$$

γ = smoothing constant to be determined by grid search

Let $\bar{E}_0 = 0$

$\text{ABS}(\bar{E}_t)$ = absolute value of the smoothed error

$\overline{\text{ABS}(\bar{E}_t)}$ = smoothed value of the absolute value of the smoothed error

$$\overline{\text{ABS}(\bar{E}_t)} = (1-\gamma)\overline{\text{ABS}(\bar{E}_{t-1})} + \gamma(\text{ABS}(E_{t-1})) \tag{3.4}$$

Let $\overline{\text{ABS}(\bar{E}_0)} = 0$

$$\alpha_t = \bar{E}_t / \overline{\text{ABS}(\bar{E}_t)} \tag{3.5}$$

α_t = the smoothing constant used in equation 3.1.

In this technique α_t is continuously varying as each new data point is received. γ is chosen by a grid search which finds the value of γ that results in the minimum mean squared error for the series X_t. Generally, γ should be less than .5. As this technique is usually not very sensitive with respect to γ, only a few values need to be tried. One should note that this technique ensures that the magnitude of α_t is always less than or equal to one.

These three adaptive forecasting techniques work by applying time-varying values to the smoothing constant in a simple exponential smoothing model. These methods are more powerful than simple exponential smoothing with fixed coefficient values, but also are more costly. As not much work has been done in the area of adaptive forecasting methods, this field contains many areas that are ripe for investigation.

Non-adaptive exponential smoothing techniques are usually simpler to implement and to understand than adaptive techniques. They are capable of excellent predictions when the underlying processes are well-defined and stable.

Following is a non-adaptive technique that makes use of exponentially weighted moving averages (i.e. exponential smoothing). It operates by separately estimating for each time t the smoothed process average, S_t, the process trend, R_t, and the seasonal factor, F_t, and combines them to compute a forecast, Y_{t+T}, T time units into the future. These values are recursive in nature and thus, each forecast value is based on all of the data from all of the times preceding and including t.

$$S_t = (\alpha)\frac{x_t}{F_{t-L}} + (1-\alpha)(S_{t-1} + R_{t-1}) \qquad 0\leq\alpha\leq 1 \tag{3.6}$$

$$R_t = (\beta)(S_t - S_{t-1}) + (1-\beta)R_{t-1} \qquad 0\leq\beta\leq 1 \tag{3.7}$$

$$F_t = (\gamma)\frac{x_t}{S_t} + (1-\gamma)F_{t-L} \qquad 0\leq\gamma\leq 1 \tag{3.8}$$

$$Y_{t+T} = (S_t + (T)R_t)F_{t-L+T} \tag{3.9}$$

where α, β and γ are smoothing constants which help to dampen sharp peaks and troughs, L is the seasonal period and x_t is the true process value at time t. If the process is on a month to month basis, then L = 12 and T = 1 and the forecast value is

$$Y_{t+1} = (S_t+R_t)F_{t-11} \tag{3.10}$$

In the discussion that follows, we will assume this to be the case.

Exogenous variables for the model are the values of S_0 and R_0 and twelve initial values for F, denoted Fi_j (j = 1, ... , 12), and are computed as follows:

(1) S_0, the initial value of the process average, is equal to x_1, the first process value.

(2) R_0, an approximation of the initial trend, is found by plotting the first twenty-four data points and estimating the trend by a straight line as shown for North Atlantic travel data in Figure 3.2. R_0 is the slope of that line.

(3) Fi_j (j = 1, ... , 12), the first cycle seasonal factors, are computed by taking the ratio of the true process value to the trend approximation value, both of which are available from the graph in Figure 3.2 as given in the equation:

$$Fi_j = \frac{\text{actual value month } j}{\text{trend approximation month } j} \qquad j = 1, \ldots, 12$$

In order to properly forecast the first several values, it is necessary to build up a backlog of values of S, R and F. This process also helps subdue the random factors inherent in the data. In practice, two cycles of S, R and F values have been found sufficient for these purposes. Thus, the first forecast value, Y_{25}, would be based, in an exponentially decreasing manner, on all the S, R and F values from t = 1 to t = 24. Then, S_{25}, R_{25} and Y_{26} are calculated,

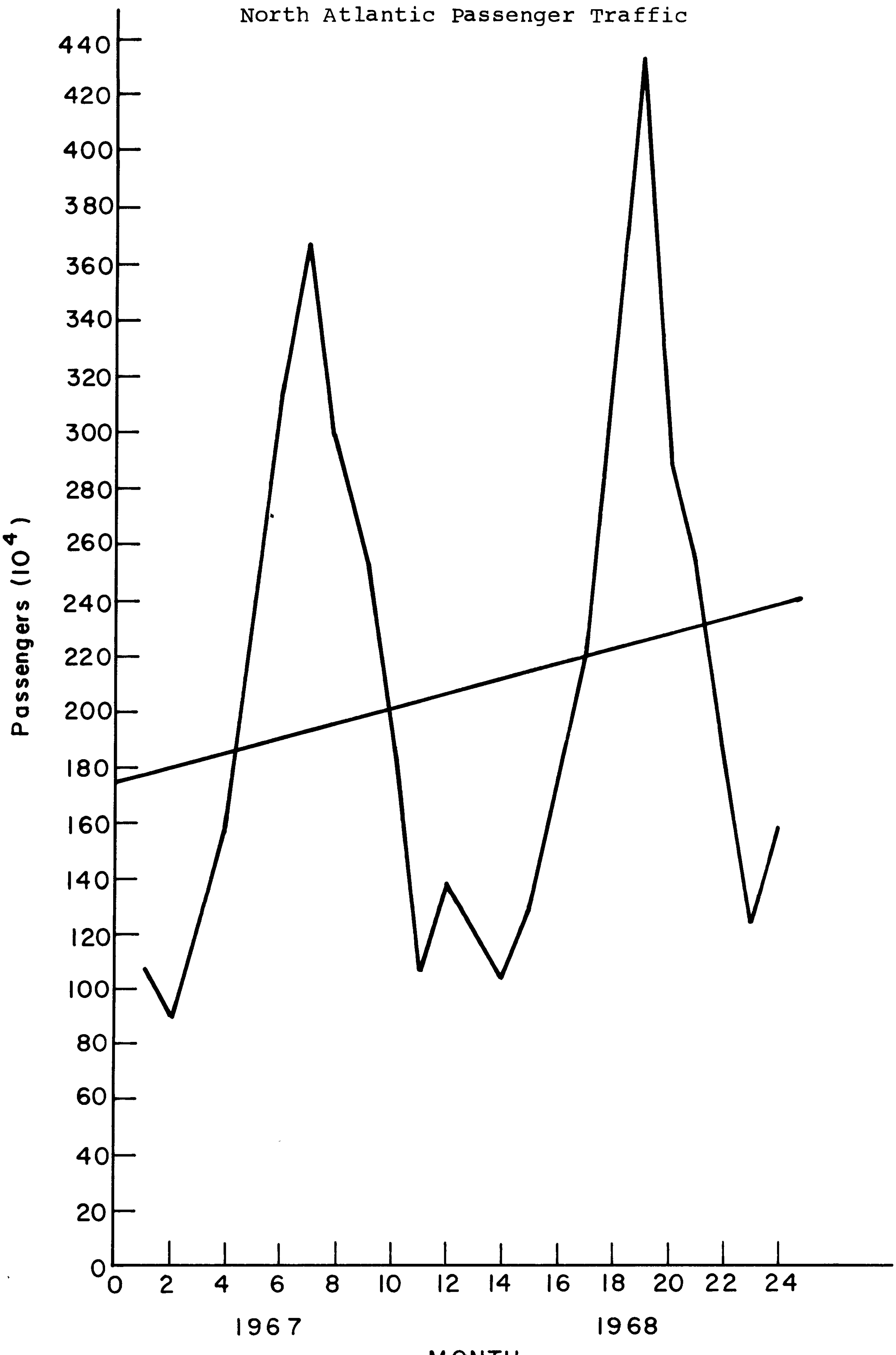

North Atlantic Passenger Traffic
440
420
400
380
360
340
320
300
280
260
240
220
200
180
160
140
120
100
80
60
40
20
0
Passengers (10^4)
0
2
4
6
8
10
12
14
16
18
20
22
24
1967
1968
MONTH

Figure 3.2

and so on. As the process continues, each forecast value is based on history in the same way.

Program Description

The program to execute this model for five years of data - two years to build up the backlog of values and three years for actual forecasting - is shown in Appendix A. As often as possible, variable names are either consistent with those in the formulae or self-explanatory. The program includes a routine that locates the combination of α, β, and γ to the nearer .01 that produces the least total squared error in the forecast values. It does this by first finding the α, β and γ to the nearer .05 and then searching around these values to see if the total squared error can be lowered further.

The program requires three sets of input data:

(1) A title card describing the forecast to be run. If no title is desired, a blank card must be inserted.

(2) The data, formatted according to statement 101 (Appendix A).

(3) One card formatted as follows:

FIRST_POINT_xxxxx_FINAL_POINT_xxxxx_YEAR_xxxxx_UNIT_x

The first point is the point of the graph (Figure 3.2) where the trend approximation line intersects t = 1. The final point is the point on the graph where the trend approximation intersects t = 24. The year is the first calendar year for which data is give; if this is unknown or inappropriate, the number 0001 whould be entered. The unit figure is a scaling factor and indicates the power of 10 by which the volumes must be multiplied.

Upon execution, one table and two graphs are printed, the former a listing of the known values, forecast values and error calculations for each time t and the latter, plots of actual and forecast volume vs. time and the error probability distribution to give an idea of the overall effectiveness of the model.

This model is applicable chiefly for short term forecasting. It is an extrapolation of history into the future. Effects of dramatic changes in technology, such as inauguration of SST service in the example above, cannot be predicted. Similarly, alterations in traffic patterns due to uncontrollable circumstances, such as strikes, cannot be predicted. Both of these instances greatly increase the random factor in the data and render the model ineffective. The model is, however, helpful over the short term and should be used with this in mind.

Box-Jenkins Time-Series Analysis

The following material is based primarily on the work of Box and Jenkins. Their technique requires a large amount of data and computational effort relative to exponential smoothing, but somewhat less than regression methods. This technique adjusts the forecasts to fit new data, and is therefore an adaptive method.

Box-Jenkins analysis assumes that one is working with a stationary series or at least a homogenous, non-stationary series. A stationary series is one in which the mean and variance remain roughly unchanging over time. A homogenous, non-stationary series is one that can be differenced one or more times, in order to arrive at a stationary series. Unfortunately, not all series are stationary or homogenous, non-stationary. It is a common mistake to apply Box-Jenkins analysis to series that are non-stationary and non-homogenous. There is not a clear dividing line between statistically suitable series and unsuitable ones. It is often a matter of judgment or necessity when one decides to apply Box-Jenkins.

In order to determine stationarity, one needs to calculate the autocorrelation functions of the series. The autocorrelation function is the covariance of the values of a series with their lagged (or lead) values. The formula for computing the autocorrelation function of a series for a lag of k periods is presented in equation 3.11.

$$\hat{\rho}_k = \frac{\sum_{t=1}^{T-k} (Y_t - \bar{Y})(Y_{t+k} - \bar{Y})}{\sum_{t=1}^{T} (Y_t - \bar{Y})^2} \qquad (3.11)$$

$\hat{\rho}_k$ = sample autocorrelation function for lag k

T = number of observations in the series

Y_t = series on which the sample autocorrelation function is computed

$\bar{Y}$ = mean of the series, Y_t

The correlogram is a plot of $\hat{\rho}_k$ versus k for k=1 to the desired number of lags). For a stationary series (as in Figure 3.3), the correlogram will quickly fall to zero.

In order to apply this method, one usually needs to transform the series, Y_t, by differencing to form a new series, w_t, which is stationary:

e.g. first differencing: $w_t = \nabla Y_t = Y_t - Y_{t-1}$

second differencing: $w_t = \nabla^2 Y_t = \nabla Y_t - \nabla Y_{t-1}$

If one has a forecast for w_t $(= \nabla Y_t)$, then in order to forecast Y_t, he has to <u>integrate</u> (undifference) the series, i.e. $Y_t = Y_{t-1} + \nabla Y_t$.

The Box-Jenkins technique makes use of ARIMA (auto-regressive integrated moving average) models: the integration (or undifferencing)

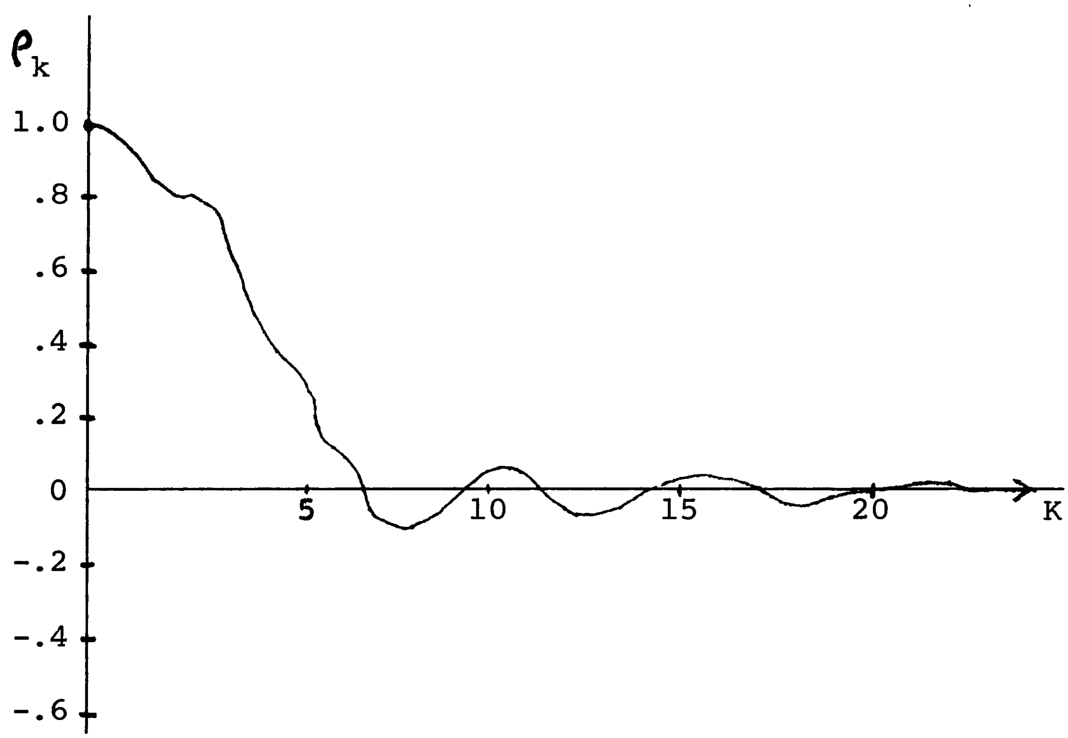


Figure 3.3. Correlogram of a stationary series

procedure has just been described, and following are descriptions of moving average, autoregressive, and complete ARIMA models. As the defining equations are straight-forward, additional discussion will be limited.

A moving average process of order q on the series, Y_t, is defined as:

$$\text{MA}(q): \quad Y_t = \mu + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \dots - \theta_q e_{t-q} \qquad (3.12)$$

μ = mean value of Y_t

θ_i = parameters associated with the moving average procedure

e_t = prediction error of the moving average process at time, t

θ_i is expected to become smaller as i increases because past values are not expected to affect the present value very much for a stationary series (ignoring seasonality considerations).

An auto-regressive process of order p on the series, Y_t, is defined as:

$$\text{AR}(p): \quad Y_t = \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \dots + \varphi_p Y_{t-p} + \delta + \epsilon_t \qquad (3.13)$$

φ_i = parameters associated with the auto-regressive process

ϵ = error associated with the auto-regressive process

δ = constant parameter needed to define the mean of Y_t (μ)

$$\mu = \frac{\delta}{1 - \varphi_1 - \varphi_2 - \dots - \varphi_p} \qquad (3.14)$$

As is the case with θ_i, φ_i is expected to become smaller as i increases.

The combination of integration, moving averages, and autoregression results in the ARIMA model.

$$\text{ARIMA }(p,d,q):\quad w_t = \varphi_1 w_{t-1} + \varphi_2 w_{t-2} + \cdots + \varphi_p w_{t-p} + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \cdots \theta_q e_{t-q} + e_t \qquad (3.15)$$

where $w_t = \nabla^d Y_t$

When applying an ARIMA model, one first has to determine the proper value of d to use in the model. This can be determined by plotting the correlogram for the series with several candidate values of d. The correlogram, practically speaking, will never die down to zero, so some threshold needs to be specified. If the magnitude of the correlogram dies down to .05 or less and stays there, then it is usually suitable. If the correlogram does not diminish to a low value rapidly (not more than approximately twenty-five lags for monthly data) then that order of d is not suitable. If several values of d seem suitable after examining their correlograms, then one might wish to try all of them in ARIMA models. All other things being equal, one should pick the smallest value of d for which the correlogram is well-behaved. As the value of d increases, one is more likely to be fitting noise and statistical fluctuations than to be fitting the true process.* Experience with numerous economic series has shown that most processes that one is likely to encounter tend to be ill-defined with only marginally acceptable correlograms.

It is unlikely that one has strong a priori reasons to pick particular values of p and q. One should try several values of p and q starting with the lowest orders. The following statistics are

*Each differencing process also destroys a degree of freedom.

available to help evaluate the models: R^2, t, and χ^2 (probability that the residuals are not white noise). One should then simulate the potentially acceptable models and observe how well they track themselves and how robust (responsive) they are.

One estimates the φ_i and θ_i by choosing the combination that results in the minimum mean squared error. As the process for estimating the parameters is highly non-linear, it is necessary to employ an iterative solution technique. These iterative solution techniques are not guaranteed to converge, although practical experience has shown that they usually do. The non-linear techniques of estimation can, and often do, result in local rather than global optima. That makes them highly dependent on initial conditions. Unless care is taken, and several different initial conditions are tried, erroneous estimates are likely to result.

Our experiences have shown that one is unlikely to be able to obtain optimal estimates for the parameters if p and q (particularly q) are greater than two. In some cases when p and q were greater than two, intuitive guesses of the parameter values gave better results than the computer estimation routine.[10] The unacceptibility of Box-Jenkins for high order techniques has been demonstrated. However, many processes need to be specified as higher order processes, e.g. to account for seasonality, monthly traffic data needs to be modelled with at least a twelfth order lag. Many time series also have a two to four year business cycle underlying them. Such cycles are virtually impossible to predict using Box-Jenkins.

Box-Jenkins works admirably for some time series, but is insufficient for others. Fortunately, another technique is available that greatly increases the power of the basic Box-Jenkins method. This method is to use economic variables (which can take out seasonal, business cycle, and other long term effects) in addition to the usual Box-Jenkins variables. Routines have been developed that allow this. This technique is roughly equivalent to performing an ordinary least squares regression with the economic variables and then performing Box-Jenkins on the residuals (errors).

Box-Jenkins analysis was attempted on the Federal Reserve Board Index of Industrial Production of Machinery (in constant dollars). It was difficult to achieve a stationary series, even with multiple differencing. When a marginally acceptable series was arrived at, the Box-Jenkins method still gave poor results. When the machinery index was regressed against the anticipations of business expenditures for new plant and equipment (in constant dollars), approximately ninety-four percent of the variation in the series was explained (including business cycle and seasonal effects). A correlogram of the residuals (errors) showed well-behaved, stationary series for several values of d. When Box-Jenkins analysis was applied to the resulting series, excellent results were obtained.

There are many other cases where simple economic indicators coupled with Box-Jenkins have yielded good results when the basic Box-Jenkins method was unsuitable. Box-Jenkins analysis is limited to low-order processes. Despite this limitation, Box-Jenkins with economic indicators is a useful tool for short term analysis.

Spectral Analysis

Spectral Analysis is a technique of evaluating a time series with time as the only predictor. Any stationary series (i.e. one with unchanging mean and variance, see Figure 3.1) can be decomposed into a series of sine waves of various frequencies (i.e. periodicities) and phase. Given enough sine waves, any series can be accurately fit. However, this may then result in the problem of fitting n points with an nth degree polynomical. In spectral analysis, one should attempt to detect cycles that really exist rather than ones that just happen to fit the series; typical series which commonly appear are the long run business cycle, seasonal peaks, and day of week cycles.

When performing spectral analysis, one should ensure a stationary series by generalized differencing or logarithmic transformations. A Fourier series can be used to break the series into a set of sine and cosine waves of varying amplitude or alternatively a set of sine waves with appropriate phase lags. With enough decomposition, any series can be fit, but a good fit does not guarantee a good prediction.

The longest period that can be observed is equal to the number of points in the sample. The shortest period that can be observed is twice the measured increment. If the mean of a stationary series is removed so that it fluctuates about zero, one can write:

$$Y_t = \sum_{i=1}^{N} (\alpha_i \cos w_i t + \beta_i \sin w_i t) \tag{3.16}$$

Y_t = stationary series to be analyzed after the mean is removed

t = time

N = number of frequencies considered

α_i, β_i = parameters to be estimated

$_i$ = frequency values (radians/sec)*

This formulation assumes that power is concentrated at discrete frequencies and that the individual cycles are independent.

The correlogram (which was described in detail in the section on Box-Jenkins models) is used in Spectral Analysis. The auto-covariances

---

*Frequency in radians per second can be converted to cycles per second by dividing by 2π.

from the correlogram are used to calculate the power density spectrum (of the lagged variables). The power density spectrum is a list of component frequencies of a series with the magnitude (or power density) associated with them. For example, a series of air passenger demand might have dominant magnitudes at frequencies corresponding to zero months (mean of the series) and twelve months (the yearly cycle). All other frequencies will probably have much smaller magnitudes (power) than those spread across some frequency range, and hence the name power density spectrum. Rather than going into details, a computer routine is presented in Appendix B which will calculate the desired quantities.

Although discrete frequencies are assumed, statistical perturbations are usually such that the exact frequency lies between two of the specified frequencies. A histogram which groups a range of frequencies together will quantize the frequencies for practical usages. The histogram is implemented through "Parzen windows" in the spectral estimation program of Appendix B. A "Parzen window" is just one of a number of complicated mathematical forms that have historically been denoted as windows. One needs to work with frequency bands rather than frequencies, e.g. the frequencies .43200001 and .4320 are essentially the same frequency and one would not want to include both in the model. There is some interval such that two frequencies are not considered as being the same. This interval is referred to as a window. The Parzen windows act as smoothers of the estimates which enable the user to observe "true" process values with statistical perturbations and random noise removed.

The spectral estimation program calculates consistent estimates of the power density spectrum* which converge to the correct value

*The power density spectrum consists of the root mean square values of amplitude at each frequency.

for large populations. There is a problem with determining the best number of lags to use. A large number of lags results in small variances; a small number of lags results in small biases. As the mean squared error is equal to the variance plus the bias squared, it is not clear what is the best number of lags to use for a given problem. Several values should be tried in order to pick the optimum number of periodicities, e.g. if one were dealing with monthly travel data on the North Atlantic, he would have to include at least twelve lags in order to account for seasonality. As the number of lags increases, there is the increasing danger of picking up noise rather than underlying patterns.

One needs to ensure that the frequency spacing is close enough that good results can be obtained. The frequency spacing interval generally should be at least twice as small as the interval between observations. Although some confidence measures can be created, there are not many statistical validity tests that can be applied to spectral analysis. One can test for removal of trend by observing the size of the zero frequency term (which should be small).

Chapter IV

Control Theory Models

This section will cover models that rely strongly on feedback mechanisms in their structure. Feedback mechanisms refer to the relationship of the output or response of a system (or a model of a system) to the system itself. For instance, a profitable year for an airline might result in the ability to increase the level of service and to purchase additional plant and equipment. This output of the "airline system" affects the system in the next time period by possibly causing a further increase in profitability. This process of system outputs affecting system inputs is known as a feedback mechanism.

The process of feedback often results in models of transportation demand being self-fulfilling. If a model predicts an increase in demand, supply often will be increased which improves the level of service, thereby causing the predicted increase. On the other hand, if large increases in demand are not predicted, new supply facilities will not be introduced and demand will, as predicted, not increase beyond the constraints of the existing supply. It is almost a truism that a shift in supply will cause a shift in demand. If the demand shift occurs slowly and smoothly over time, it is not necessary to model feedback effects. If this is not true, there might be much to be gained by modelling the interactions of various parts of a transporation system with feedback relationships.

Information filtering will also be discussed here as these techniques developed from the same branch of engineering as the techniques of analyzing and designing feedback control systems. Information filtering is the process of extracting useful information from a set of observations. Most observations will consist of the true process value plus noise, plus possibly a bias. An information filter might attempt to estimate the bias and to cancel out the effects of noise. A very simple means of filtering might be used to compare the passenger appeal of two different aircraft types. Each passenger can be asked to rank the aircraft on a numerical scale. Averaging their answers will presumably cause the random factors to cancel out and result in an unbiased estimate of the true value. An analogy can be made to understanding a conversation in a noisy room. The listener's mind will tend to reject the noise and hear only the voice (true signal) thereby acting as an information filter. The theory behind the analysis of feedback models and information filtering is a highly detailed and mathematical subject. A very brief and by no means complete overview of the techniques will be presented; to go into more detail would require a very lengthy digression which is out of the context of this report and is not necessary for the application of these methods. It is advised that the reader who is interested in a more rigorous discussion consult Forrester, Principles of Systems; Savant, Basic Feedback Control System Design; Battin, Astronautical Guidance; or any other book that deals with feedback systems and information filtering.

A system may be modelled as an open or closed loop. An open loop system is one in which there is no interaction of the outputs with the inputs. An open loop system is not aware of its own behavior. A firm in which a fixed amount of funds is allocated to advertising and promotional activities regardless of what results occur is acting as an open loop system. If the results are predictable and stable, this procedure might be acceptable. A closed loop (or feedback) system is influenced by its past behavior. Outputs at a given point in time affect future inputs. If in the case of an advertising budget, management reviewed its budget based on previous results and present goals, the system would be closed loop.

Over the short term, a fixed schedule airline might be thought of as an open loop system in that a flight will depart whether only one passenger is aboard or the flight is over-booked. The Eastern Airlines Shuttle is a closed loop system. Departures are demand responsive as the number of arriving passengers feeds back to the scheduler, and the appropriate number of flights are dispatched. Most air transportation models are open loop and pay little attention to dynamics. The dynamic or time-varying behavior of a system can be important for some purposes. One must be aware that changes in fare or level of service will affect passenger demand over a period of time, not instantaneously. For most purposes, dynamics and feedback can be safely ignored. However, over the long term, everything is ultimately a closed loop system. For long term analysis and forecasting, consideration of these feedback effects must be included either explicitly in a control theory or simulation model, or implicitly through the modeller's intuitive mental model.

Some feedback relationships are positive. In positive feedback loops, a given response results in actions that cause further response of the same nature. Such behavior often manifests itself as exponential growth or decay. As more aircraft are built, more are bought and flown. This hopefully causes an increase in safety, comfort, convenience, and frequency; thereby causing an increase in passengers. This increase in passengers further triggers more research and development which results in newer and better aircraft; thus completing the loop.

Feedback relationships can also be negative. Negative feedback loops can be thought of as "goal seeking." Their response tends to direct them toward a goal that is either recognized or implicit. A goal might be to achieve ninety percent of departures within five minutes of schedule, to achieve a specified load factor (usually constrained by the system rather than by management), or some other undefined achievement. As the number of passengers increases, load factors tend to increase, which causes availability and comfort to decrease while processing time increases. This decrease in level of service tends to decrease the number or passengers and results in a self-regulating system.

These examples are overly simplistic. For instance, in the last example, an increase in passengers might cause an increase in frequency which would result in an increase (not a decrease) in the level of service. Systems usually contain several feedback loops which operate simultaneously. At various times, different loops can potentially dominate the system's response. Loops can operate at various levels of detail. In the previously described negative feedback loop, level of service decrease can be linearly related to load factor, or by simulation involving a complex generation technique for reservation requests and a mathematical model of processing delay, each with its own feedback loops.

An individual flight, a city-pair, an airline, a complete multimodal transportation system, cash flows in a company, or the entire economy can be modelled. As more loops become present, analysis and intuitive understanding dramatically increase in difficulty. As in other types of analysis, it is desirable to eliminate all unnecessary detail.

A system can be defined as a function of the state (or condition) of some quantity by simultaneous differential equations. Such precise and mathematical relationships tend to be of marginal value relative to less exact techniques because even if all specification errors were eliminated and the model were perfect, measurement errors and/or inability to make use of the increased accuracy negate any benefits derived from such complicated methods. Only the simplest of systems can yield a closed-form solution* that can be solved as a function of time and exogenous variables. More complicated systems require simulation or other advanced techniques for solution.

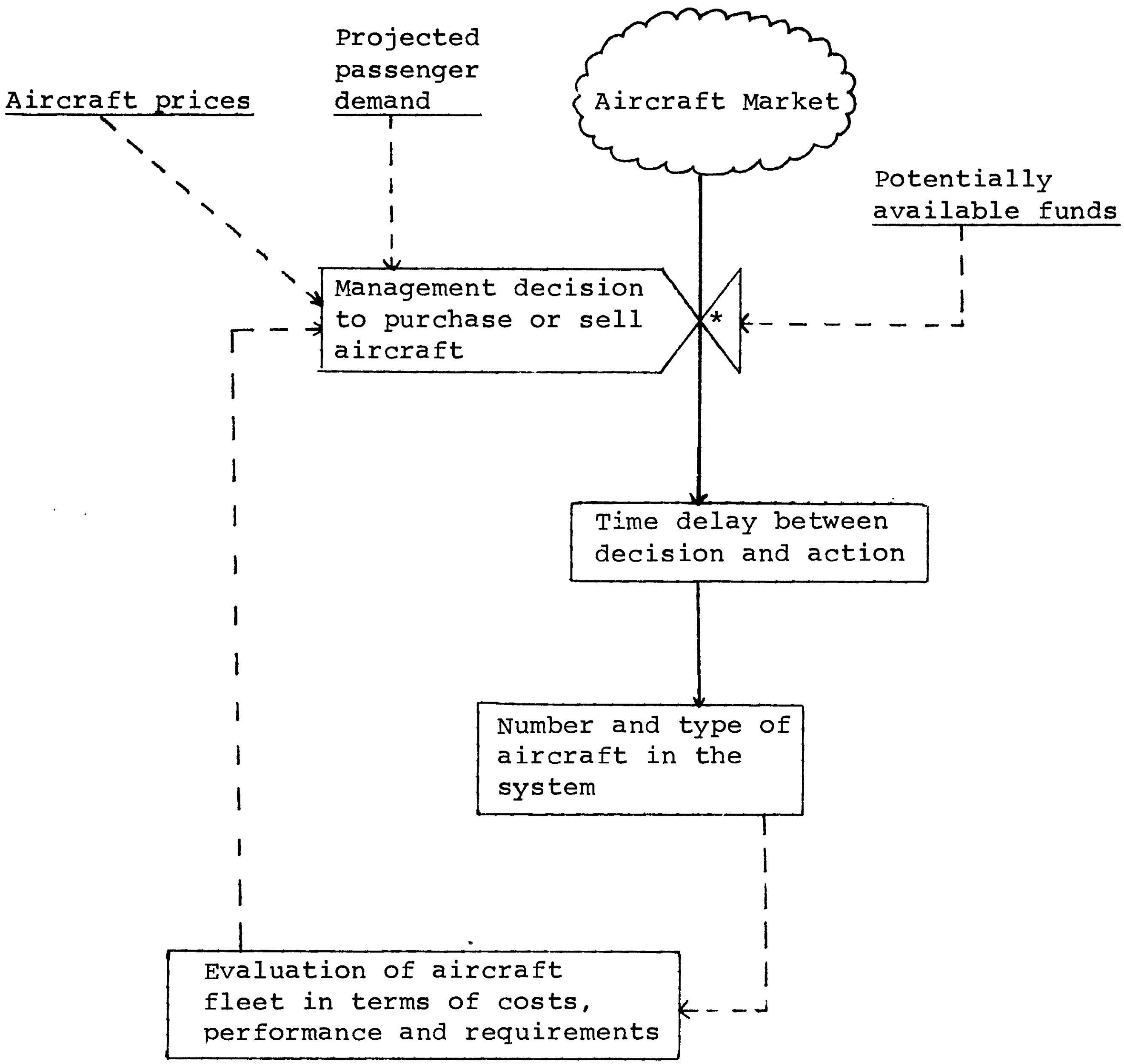
An example of a simple feedback loop is shown in Figure 4.1. Although this loop appears to be quite simple, any part of it might require a complex econometric regression to define it. Negative loops can be characterized as self-regulating or "goal-seeking." Positive loops can be characterized by growth or shrinkage. Most systems are difficult to define as either one or the other although over certain ranges of values, it might be possible to describe the actions of a system by one or two predominate loops.

Systems that are non-linear or which are higher order (second derivatives or higher) can demonstrate oscillatory behavior, as in Figure 4.2. This could be a simple response such as a product life cycle, or it could be a more complicated response such as might occur when a new, highly attractive, limited availability fare is introduced; demand for the new fare could develop very quickly, and potential demand could overshoot supply, thus causing a decrease in demand; this process could continue until the oscillations dampen out.

Techniques for solving linear and certain non-linear control problems can be used for some closed-loop transportation systems. These techniques require familiarity with differential equations, LaPlace transforms, and other engineering tools. One

---

*A closed-form solution is one with a unique and exact answer that can be arrived at through simple algebraic operations without the need of applying iterative or other approximate procedures.



Note: Solid lines represent flows of physical quantities.
Dotted lines represent flows of information.

Figure 4.1. Basic Feedback Loop

*This symbol represents a valve or switch through which changes and actions occur.

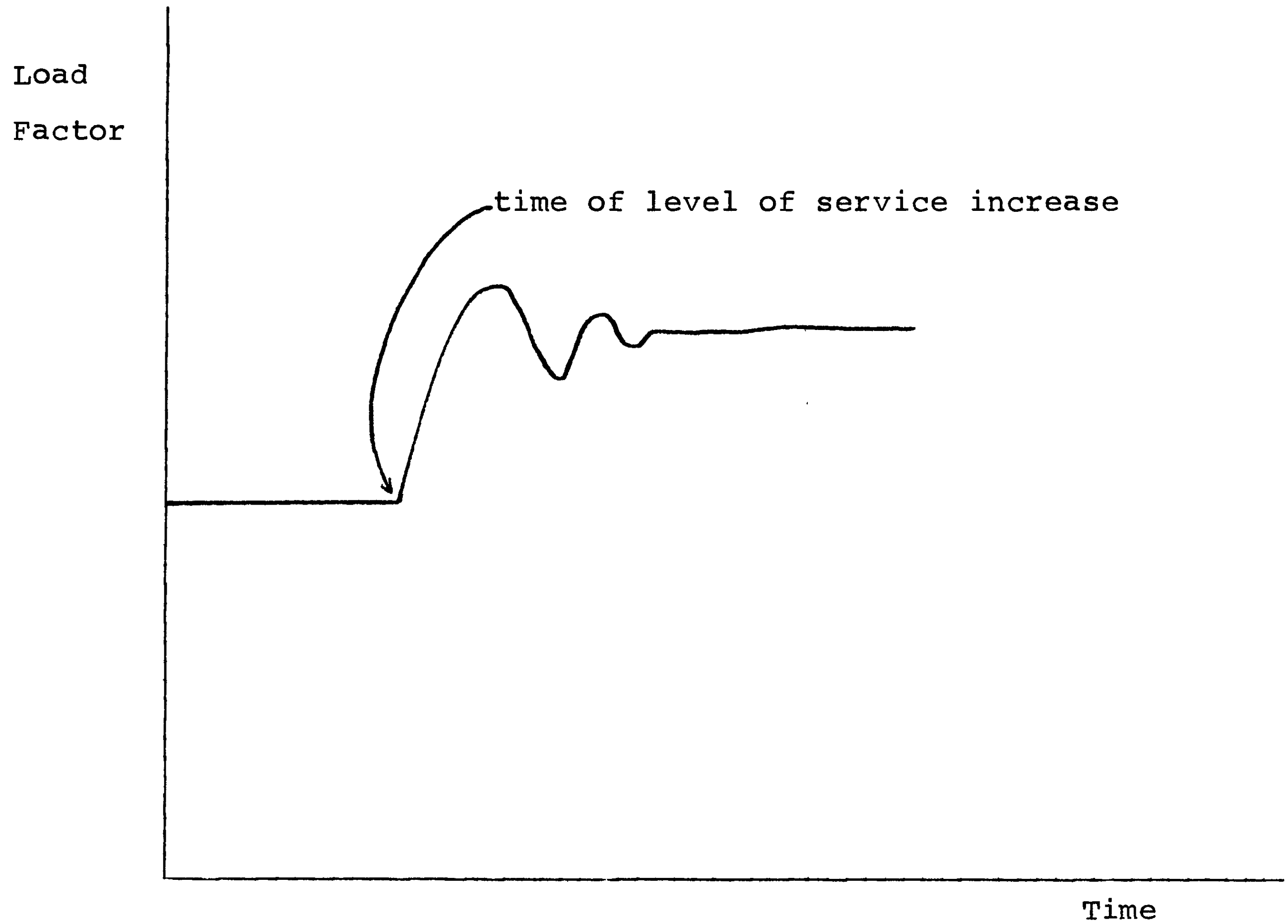


Figure 4.2. Oscillatory response of seasonally adjusted load factor to an increase in level of service

need not concern himself with these methods as they are unlikely to ever be explicitly needed by transporation planners. It is also possible to simulate simple systems on an analog computer. More complicated systems can sometimes be modelled on a hybrid computer.*

The first step for analyzing a feedback system is to break the system into various blocks or black boxes each of which has a specific function that can be characterized by a transfer function (mathematical expression) that relates inputs to outputs. The model can be best understood if the blocks are transparent so that the inner workings can be observed if desired. There exist many modern and classical techniques for analysis of feedback systems.

Feedback models can be initialized at a given state and then be left to run freely, thus giving a time series picture of the system. It is far more valuable to allow the system to respond to exogenous inputs through appropriate behavioral assumptions. A model of passenger demand could have the capability of responding to factors such as personal income, characteristics of other modes, fare, and level of service changes. A very simple model of such a process is shown in Figure 4.3.

Even state-of-the-art techniques of control theory are limited to linear or special non-linear systems. Most problems in transportation demand forecasting tend to be more complicated. System dynamics, which is a less technical, user-oriented technique derived from control theory, gets around this problem by recursively simulating the system for small time steps.

In system dynamics, a system is broken into levels and rates. Levels are analogous to states (or condition of a system), and rates

---

*An analog computer consists of a set of electrical components that can replicate the actions of a system. A hybrid computer is a combination of an analog computer and a regular digital computer.

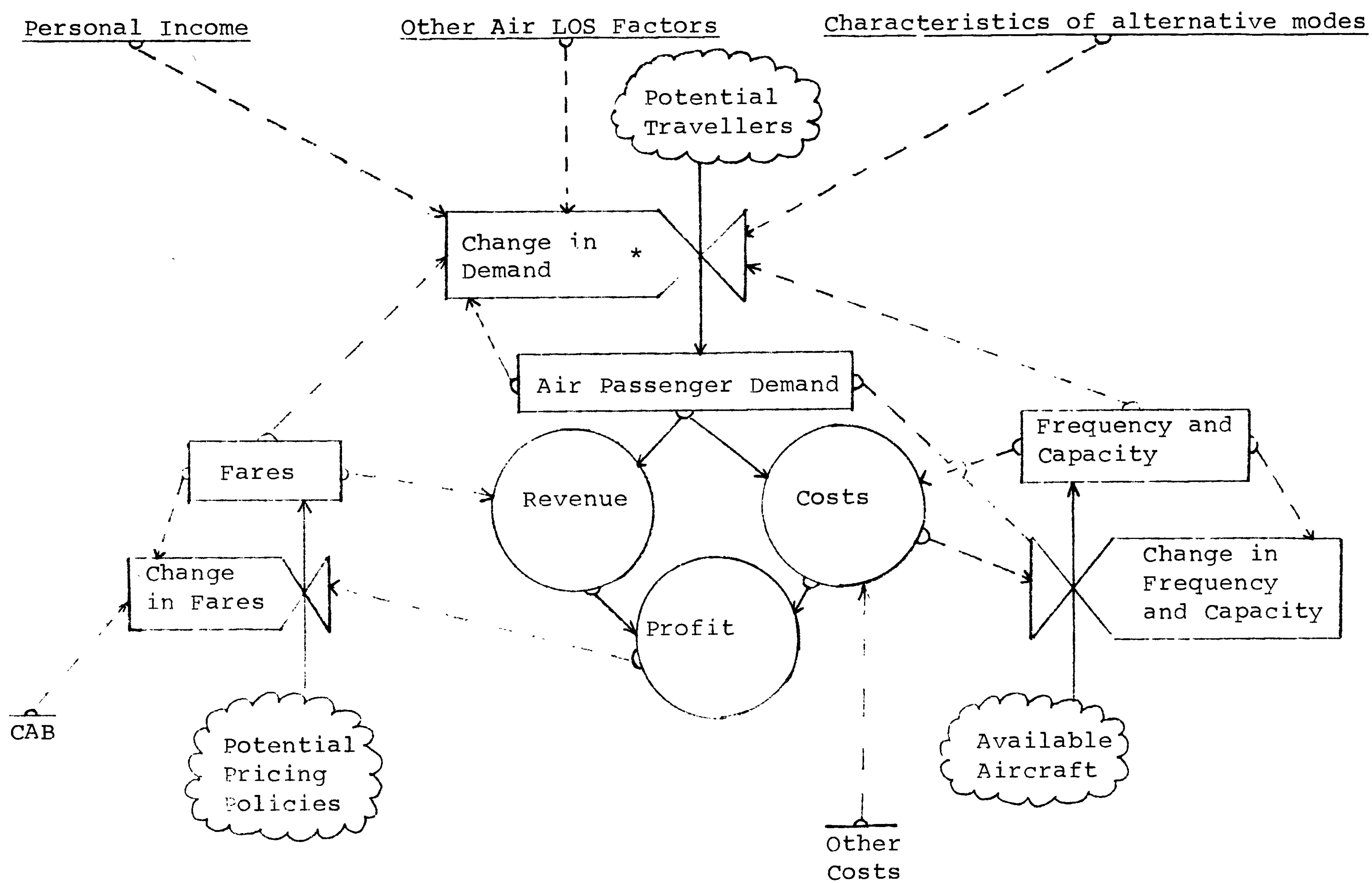


Figure 4.3. A Simple Model of Air Passenger Demand

* The change in demand module will probably consist of a short-run econometric demand equation

correspond to changes in states. Levels represent integration (accumulation) over time of the rates. The emphasis of system dynamics on integrations does not represent a real mathematical difference between the differential equation approach of control theory. It is probably more convenient for most people to think in terms of integrations, e.g. passenger loads are integrations over time of reservation requests rates. Figure 4.4 illustrates the structure of a model that can be used for evaluating frequency changes on a route. This should illustrate the basic technique of setting up a model using levels and rates.

System dynamics or control theory models often require inputs from econometric models to create transfer functions or to relate levels and rates. To the uncertainty of the usually short run econometric parameters is added the much greater uncertainty of the long and short run feedback relationships. There is both the danger of ignoring variables and relationships that might become important at some later date and the problem of misspecification errors growing and possibly becoming explosive as the feedback process is simulated through time. One should be very wary of unstable systems; real world transportation systems do not usually exhibit unstable response.

For previously mentioned reasons, control theory and system dynamics models are likely to be poorly suited for making accurate predictions. For short or medium term forecasting, econometric or simulation models are more useful and more accurate for many problems. The greatest value of feedback models is in their revealing of general system response. Much greater confidence can be placed in a curve derived from a feedback model than in one simply hypothesized as an S-shaped curve. Quite often, feedback models applied to other areas in the social and physical sciences have

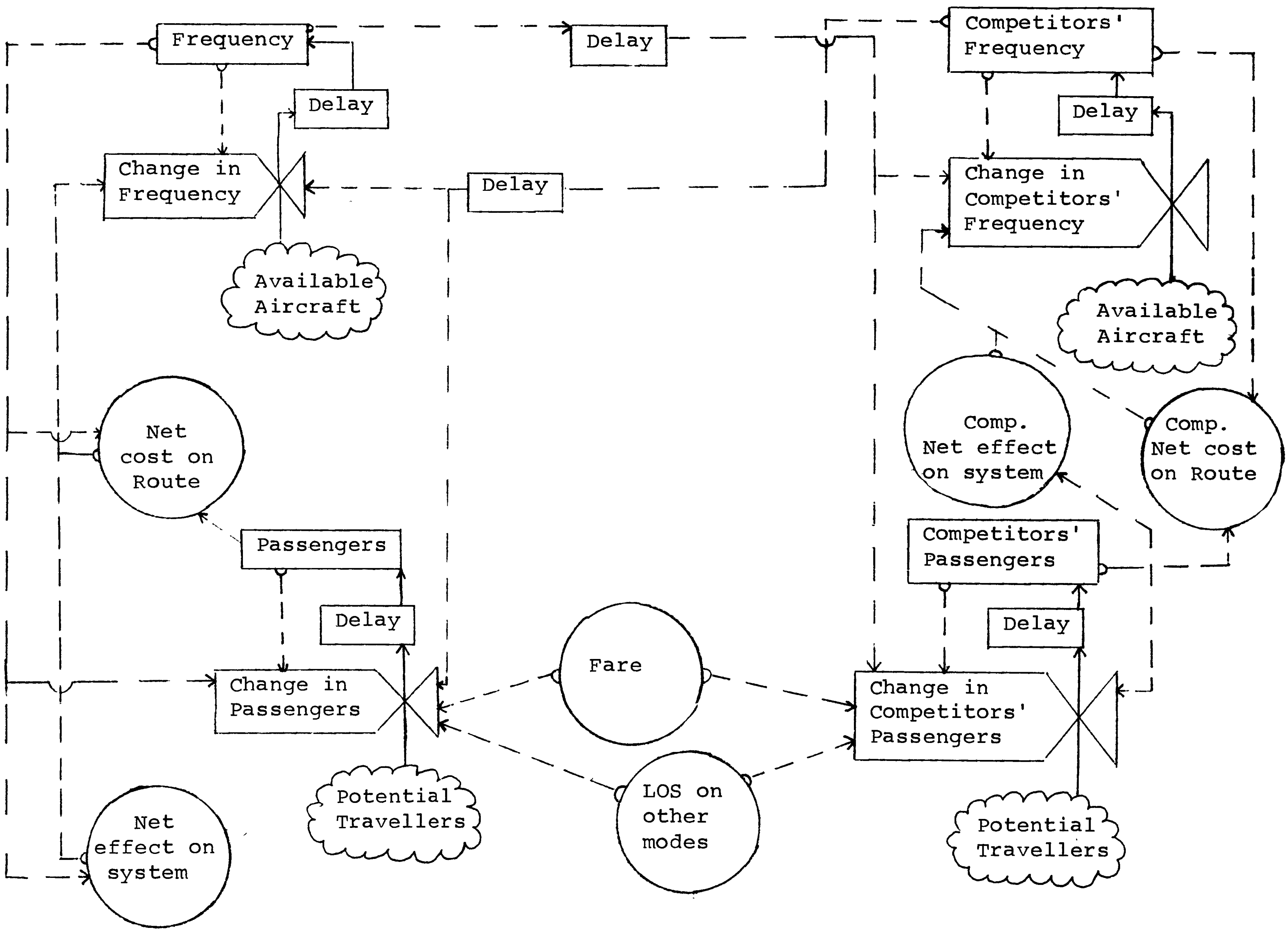


Figure 4.4. Simplified System Dynamics Model of Decision to Change Frequency

yielded counter-intuitive results. There is no reason to think that this would not be true for problems in air transportation. Even a person with a good intuitive understanding of a situation might have difficulty predicting traffic when the dominant structure changes. A good feedback model and analysis of the general shape of the response can do much to help improve one's knowledge of the macroscopic nature of a system.

Information filtering serves an important function in aircraft navigation and other physical problems. This technique can be modified to improve forecasting methods. Information filtering schemes extract useful information from processes that have inexact or noisy outputs. The discussion and example following should make the concept of an information filter clearer. The Kalman filter is a relatively simple and effective information filter that has been used successfully for aeronautical problems. Following will be a non-technical overview of Kalman filtering and suggestions for its application to air transporation problems.

Any system can be characterized by equations that relate the dependent quantities from one period to the next. Measurements are available from which the dependent quantities can be calculated. There will always be errors that are associated with the measurements and the model itself.

The Kalman filter is a set of mathematical relationships that processes the available measurements optimally (i.e. with the minimum mean squared error) in order to achieve the best possible estimate of the dependent quantities. The filter is simply a sophisticated method of exponential smoothing that can estimate bias errors and can cancel out much of the random errors. The basic filtering algorithm consists of forecasting future values of the dependent quantities, and then processing measurements with the filter in order to update the estimates of the dependent quantities to reflect the new information contained in the measurements.

As this subject tends to be highly mathematical, the reader might wish to skip the computational details that follow. The technique can be applied to almost any forecasting method without requiring an understanding of why the filter works. Simple user-oriented computer routines can be developed to solve this problem in much the same way that user-oriented regression routines allow one to perform least square analysis without knowing the details of the calculations.

In the general case, one needs to define a state and an error covariance matrix that goes with it. The state should consist of all dependent quantities that one might wish to estimate (such as passenger volume, market share, profits, etc.), measurement and model misspecification bias (note: measurement bias arises primarily from errors in forecasts of the exogenous variables), and any other quantities that are needed to extrapolate the state to future time periods. Each of the elements of the state has some error distribution associated with it, i.e. its value is not known exactly. The errors can usually be assumed to be independently distributed, and can be characterized by variances (VAR(x)) and covariances (COV(x,y)).

$$\mathrm{COV}(X,Y) = \mathrm{COV}(Y,X) = \frac{\sum_{i=1}^{N} (X_i-\bar{X})(Y_i-\bar{Y})}{N-1} \tag{4.1}$$

$$\mathrm{VAR}(X) = \frac{\sum_{i=1}^{N} (X_i-\bar{X})^2}{N-1} \tag{4.2}$$

X, Y = any variables

$\bar{X}, \bar{Y}$ = mean values of X and Y respectively

N = number of observations of X and Y

The covariance of X and Y represents the relationship of the errors of X and Y. A covariance of zero implies that no relationship exists between the errors of X and Y. A negative covariance implies that positive errors in X are most likely to be associated with negative errors of Y. Positive covariances imply that the errors tend to move together. For instance, if one were modelling passenger demand, positive errors in the forecast of income would tend to accompany positive errors in forecasting demand.

The covariance matrix is a convenient form for relating the errors of the state to each other. The diagonal elements of the matrix are variances rather than covariances. The variances are simple to estimate because they are just the square of the standard deviation which one usually has a good estimate of. The covariance matrix contains the estimates of uncertainty in the state. To initialize the process, an initial covariance matrix, such as one where all off-diagonal terms are set equal to zero, needs to be specified. Elements in the off-diagonal terms should quickly reach reasonable values.

The state error covariance matrix ($\mathbf{E}$) and the state ($\mathbf{x}$) are extrapolated through time as forecasts are made. For some problems, the extrapolation process is well defined by physical relationships. In air transportation problems, one needs to rely on the less exact extrapolation that results from an econometric regression, a feedback model, or a simulation model. The process is not exactly defined, and one can only hypothesize reasonable simplifications. However, one should note that Kalman filtering has been successfully applied to processes that are even less exactly defined than the forecasting of the demand for air transportation.[11]

The state extrapolation is an easier process to understand than the covariance matrix extrapolation. The state extrapolation is just defined by the single equation or multi-equation system which has been hypothesized and then calibrated by regression analysis or other techniques. The expression for this is shown in equation 4.3.

$$\mathbf{x}' = \mathbf{\Phi x} \tag{4.3}$$

$\mathbf{x}$ = the state

$\mathbf{\Phi}$ = the transition matrix that relates $\mathbf{x}_{t+1}$ to $\mathbf{x}_t$.

Throughout the rest of this chapter, the prime notation (e.g. $\mathbf{x}'$) represents the extrapolated or updated value of the quantity being considered.

The transition matrix ($\mathbf{\Phi}$) specifies the relationship between the state at time (t) and time (t+1). If the relationships are known, the elements of $\mathbf{\Phi}$ can be solved by inspection after the expression in equation 4.3 is expanded into scalar notation. For constant terms, the transition matrix has a one in the appropriate diagonal elements and zeroes in the rest of the row. The following example should clarify the procedure for computing the elements of the transition matrix.

Assume the following state:

$$\mathbf{x} = (Y,Z)$$

Y = number of employees in the marketing department of a firm

Z = number of hires for the entire firm in the present time period.

Assume that the number of hires in each period is constant.

$$Z_{t+1} = Z_t \tag{4.4}$$

Also assume that Y has been defined by the following regression equation:

$$Y_{t+1} = .95\ Y_t\ t\ .2\ Z_t \tag{4.5}$$

From equation 4.3, one can write

$$\begin{bmatrix} Y_{t+1} \\ Z_{t+1} \end{bmatrix} = \begin{bmatrix} \Phi_{11} & \Phi_{12} \\ \Phi_{21} & \Phi_{22} \end{bmatrix} \begin{bmatrix} Y_t \\ Z_t \end{bmatrix} \tag{4.6}$$

expanding 4.6 into scalar notation results in equations 4.7 and 4.8.

$$Y_{t+1} = \Phi_{11}\ Y_t + \Phi_{12}\ Z_t \tag{4.7}$$

$$Z_{t+1} = \Phi_{21}\ Y_t + \Phi_{22}\ Z_t \tag{4.8}$$

From equations 4.5 and 4.7, it is clear that Φ_{11} equals .95 and Φ_{12} equals .2. From equations 4.4 and 4.8, it is obvious that Φ_{21} equals zero and Φ_{22} equals one. Therefore

$$\mathbf{\Phi} = \begin{bmatrix} .95 & .2 \\ 0 & 1 \end{bmatrix} \tag{4.9}$$

The covariance matrix is extrapolated by pre-multiplying it by the transition matrix ($\mathbf{\Phi}$) and then post-multiplying it by the transpose of the transition matrix. Process noise ($\mathbf{Q}$) is usually added to improve performance and to help insure stability. The process noise accounts for the fact that as a process is extrapolated in time uncertainty in the process increases. This uncertainty is caused by both mathematical and practical reasons. The mathematical reason is that as a value is extrapolated beyond the sample mean, the prediction interval (which indicates the degree of uncertainty in the prediction) at a given confidence level increases. The practical reason is that assumptions of linearity or simple specifications are likely to diverge from the real world or even to break down as the model is applied to increasingly unfamiliar situations. With process noise added, the prime cause of instability is computer round-off. Equation 4.10 defines the covariance matrix extrapolation procedure. Computation of the transition matrix depends on the particular model formulation being used.

$$\mathbf{E}' = \mathbf{\Phi E \Phi}^{\mathbf{T}} + \mathbf{Q} \qquad (4,10)$$

Error in the raw model comes from specification error and measurement error. In transportation systems, as opposed to physical systems, specification error tends to be large relative to measurement errors. To review, the state should include the dependent variables and all endogenous variables. It should also include a model specification bias as well as biases for each of the coefficients (or parameters depending on view point). Other qualities can be included in the state if necessary. It is probably best to model the biases as constants which do not change through extrapolation. The biases can be either percent or absolute biases; more complicated forms are not usually necessary. The biases do not necessarily represent a unique physical quantity in transportation problems. They might be some combination of measurement and specification errors. Their value therefore lies primarily in forecasting, not in the determination of structural coefficients.

In most filtering applications frequent measurements are taken that result in updating of the state and of the covariance matrix. Many air transportation systems do not have this frequent data available, e.g. North Atlantic travel is not reported at very frequent rates, therefore, it might take months and possibly years before performance transients die down and the filter becomes well-behaved. However, Kalman filtering is well suited to problems where frequent observations are possible, such as the prediction of the volume of traffic a particular airport will handle, or the demand for an individual flight or city-pair.

The observations (measurements) are used to update the state and the covariance matrix. The filter might use pseudo measurements that are manipulations of the actual measurements to update the state and the error covariance matrix (see the example following for an example of pseudo-measurements). The relevant equations are presented in 4.12, 4.13, and 4.14. To perform and update, it is necessary to form ΔQ (the difference between the predicted and actual measurement), the measurement vector ($\mathbf{b}$), and the weighting vector (ω).

The measurement vectors specify how changes in the measurements are related to changes in the state. For linear systems or small changes, the partial derivatives of the measurement with respect to the state (i.e. the slopes) indicate this relationship:

$$\mathbf{b}_q = \frac{\partial q}{\partial \mathbf{x}} \tag{4.11}$$

where q is the measurement type.

Note that the derivative of a scalar (q) with respect to a vector ($\mathbf{x}$) is a vector ($\mathbf{b}_q$) whose elements are the derivative of the scalar with respect to each of the elements of the vector ($\mathbf{x}$).

The weighting vector specifies how the state will be updated for a unit error in the measurement. The expression for the weighting vector is found in equation 4.12. To repeat, the measurement vector relates the state variables or model coefficients to the measurement. For many measurement types, many elements of the measurement vector are either zero or one because a measurement error might have no effect on an element of the state or would directly affect it if the quantity being measured were also part of the state. The form of the measurement vector depends on the problem being studied. Once again, the inexactness of the relationship results in transportation problems being less well defined than physical problems. It is emphasized that an understanding of the derivations of the filter equations are not necessary in order to be able to use it.

Following are the filter equations that allow the state and covariance matrix to be updated when new measurements are taken.

weighting vector: $$\omega = \mathbf{E}\mathbf{b}/(\mathbf{b}^T\mathbf{E}\mathbf{b} + \bar{\alpha}^2) \qquad (4.12)$$

where $\bar{\alpha}^2$ is an estimated value of the variance which is picked to insure good performance

state update: $$\mathbf{x}' = \mathbf{x} + \omega\Delta\mathbf{Q} \qquad (4.13)$$

covariance matrix update:
$$\mathbf{E}' = \mathbf{E} - \omega\mathbf{b}^T\mathbf{E} \qquad (4.14)$$

Other modifications such as square root formulations, measurement underweighting, or non-linear compensation via second partial derivatives can be used to improve performance and to help insure stability. These techniques can be found in the current literature of Kalman filtering. As long as sufficient precision is maintained

(>20 bits per mantissa for most problems*), these more advanced techniques are not necessary. This technique is particularly useful when serial correlation or multicollinearity is present. The biased coefficient estimates are corrected for by estimating the biases.

The derivations of the filter equations are beyond the scope of this report. It is not necessary to be familiar with the mathematical details in order to be able to apply Kalman filtering to transportation-related problems.

The following example will illustrate the method of Kalman filtering as applied to a simple problem. Although this particular model is so simplistic that it is unlikely to ever be used for serious forecasting, the application of Kalman filtering to more sophisticated models follows directly.

Assume the following model:

$$V = a + b*Fare + c*Income \tag{4.15}$$

V = passenger volume

a,b,c = coefficients estimated by ordinary least squares or some other regression technique

Fare and Income are in constant dollars

The volume can be extrapolated to future periods by the following equations:

$$\text{exact form:} \quad V' = V + \Delta V \tag{4.16}$$

$$\text{approximate form:} \quad \widehat{V'} = V + \widehat{\Delta V} \tag{4.17}$$

note: the prime notation (V') refers to the extrapolated value; the hat notation ($\widehat{\Delta V}$) refers

*Single precision on most IBM computers is 24 bits per mantissa.

to the estimated value.

ΔV = volume change from previous time period

$$\Delta V = V' - V = b(\text{Fare}'-\text{Fare}) + c(\text{Income}'-\text{Income}) \qquad (4.18)$$

The volume for future time periods can be estimated by equation 4.19.

$$\widehat{V'} = V + B_{spec} + b(\widehat{\Delta \text{Fare}}+B_{Fare}) + c(\widehat{\Delta \text{Income}}+B_{Inc}) \qquad (4.19)$$

$\widehat{V'}$ = predicted value of passenger volume at a future time period

V = previously measured value of passenger volume

$\widehat{\Delta \text{Fare}}$ = estimated value of Fare change (constant dollars)

$\widehat{\Delta \text{Income}}$ = estimated value of Income change (constant dollars)

B_{spec} = bias associated with specification errors

B_{Fare} = bias associated with the prediction of Fare

B_{Inc} = bias associated with the prediction of Income change

The predicted fare change and the predicted income change (both in constant dollars) will probably have random errors and biases associated with them. If one can estimate these biases by comparing past predictions to actual values, then the bias can be corrected by adding it back in as done in equation 4.19.

Even if one had perfect predictions of fare change and income change, the model would not give the correct answer due to specification errors. The specification error is modelled as a bias term and a random term. As with the income and fare biases, the specification bias is estimated and then corrected for in equation 4.19.

The state ($\mathbf{x}$) consists of $\widehat{\Delta V}$, B_{spec}, $\widehat{\Delta \text{Fare}}$, $\widehat{\Delta \text{Income}}$, B_{Fare}, and B_{Inc}. The biases will be modelled as constants that do not change during extrapolation. As previously mentioned, one can write:

$$\mathbf{x}' = \boldsymbol{\Phi}\mathbf{x} \tag{4.20}$$

from which it follows that*:

$$\begin{bmatrix} \widehat{\Delta V} \\ B_{spec} \\ \widehat{\Delta Fare} \\ \widehat{\Delta Income} \\ B_{Fare} \\ B_{Inc} \end{bmatrix}' = \overbrace{\begin{bmatrix} 0 & 1 & bZ & cY & b & c \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & Z & 0 & 0 & 0 \\ 0 & 0 & 0 & Y & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}}^{\Phi} \begin{bmatrix} \widehat{\Delta V} \\ B_{spec} \\ \widehat{\Delta Fare} \\ \widehat{\Delta Income} \\ B_{Fare} \\ B_{Inc} \end{bmatrix} \tag{4.21}$$

$$Z = \frac{\widehat{\Delta Fare}'}{\widehat{\Delta Fare}} \tag{4.22}$$

$$Y = \frac{\widehat{\Delta Income}'}{\widehat{\Delta Income}} \tag{4.23}$$

Y and Z are exogenously generated expressions for which many independent forecasts are available.

One can measure Fare, Income, and V. From these three measurements, one can create three pseudo-measurements after the fact that relate the predicted measurements from the previous period to the actual measurements. Pseudo measurements:

$$B_{Fare} = \Delta Fare - \widehat{\Delta Fare} \tag{4.24}$$

$$B_{Inc} = \Delta Income - \widehat{\Delta Income} \qquad (4.25)$$

$$B_{spec} = \Delta V - \widehat{\Delta V} \qquad (4.26)$$

The measurement vectors (**b**) for each pseudo-measurement can be calculated from equations 4.11 and 4.19. The results of this differentiation are presented in equations 4.27, 4.28, and 4.29.

$$\mathbf{b}_{B_{Fare}} = \begin{bmatrix} b \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} \qquad (4.27)$$

$$\mathbf{b}_{B_{Inc}} = \begin{bmatrix} c \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \qquad (4.28)$$

$$\mathbf{b}_{B_{spec}} = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \qquad (4.29)$$

As previously mentioned, process noise is added each time that the covariance matrix is extrapolated in order to reflect additional uncertainty that occurs from the extrapolation process. The following process noise matrix (Q) will give acceptable results.

$$Q = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & q_1 & 0 & 0 & 0 & 0 \\ 0 & 0 & q_2 & 0 & 0 & 0 \\ 0 & 0 & 0 & q_3 & 0 & 0 \\ 0 & 0 & 0 & 0 & q_4 & 0 \\ 0 & 0 & 0 & 0 & 0 & q_5 \end{bmatrix} \qquad (4.30)$$

q_i = conservative values experimentally picked to ensure good performance.

No process noise is added to $Q_{1,1}$ because q_1 picks up all of the effect of additional uncertainty for $Q_{1,1}$ and $Q_{2,2}$. The off-diagonal values are set to zero because it is not clear until after the fact how the process interacts. An initial covariance matrix needs to be specified. A suitable one consists of zeros on the off-diagonals and conservative values of variance on the diagonals.

All of the quantities necessary to operate the filter have been defined. Equations 4.3 and 4.10 are applied each time that the state is extrapolated to a new time. Equations 4.12, 4.13 and 4.14 are applied three times each time step, once for each of the pseudo-measurements. After a few updates, the filter performance transients should die down, and the estimates should be significantly better than the raw econometric prediction.

Despite this mathematical discussion, in order to use Kalman filtering, one just needs to be able to apply five equations, 4.3 4.10, 4.12 4.13, 4.14. This can be done mechanically without the need of understanding the mathematical rigor.

One might ask why go to the additional effort of applying Kalman filtering. It is true that Kalman filtering will not be better able to explain a model's structure than a raw econometric model. Its value, in this context is to improve forecasting accuracy. It is recognized that any model, no matter how well specified, will have errors in it. Although it is desirable, it is not likely that these errors will manifest themselves as random, independent values. In such cases. the errors can be better approximated by a bias error and a random error. If one can estimate the bias error in a model, he can correct for it after the initial prediction is made. The Kalman filter is a sophisticated smoothing scheme that can estimate these errors with the minimum variance in the estimates. It is an adaptive smoothing technique that picks the smoothing constants based on the expected amount of error (as predicted by the covariance matrix) that is associated with the process. If the process contains no bias errors, the filter will not estimate them and one can then revert to simply using the raw econometric model without bias estimates. As the errors vary with time. time-varying smoothing constants can be picked; very few techniques allow for time-varying changes in the model.

Skeptics might argue that although Kalman filtering can be applied to well-defined physical processes, it cannot be applied to the less well-defined problems of transportation forecasting. Research by Fagan[11] has demonstrated that Kalman filtering can be applied to processes that are even less well-defined than transportation problems. Fagan applied Kalman filtering to the prediction of stock market prices. When compared against the best state-of-the-art techniques, Fagan's simple filter reduced the standard error of the forecast significantly (approximately ten percent). A more sophisticated version of a Kalman filter would probably result in even better forecasting accuracy.

This area is just beginning to be investigated for use in transportation forecasting. Further work will be necessary before this technique can be widely and generally applied.

The techniques of control theory have not yet been exploited for air transportation passenger forecasting. Once econometric techniques have squeezed the most possible out of the data, further improvement in forecasting can come from improved data, better knowledge of the process, and techniques developed along non-econometric lines, including simulation models and control theory models. Although they present an avenue for future research and advancement, control theory models are not yet well enough developed to significantly aid most forecasters.

Chapter V

Econometric Models

Econometric methods include numerous modelling techniques which are based on many different statistical formulations. This topic alone could be the basis for an entire report on forecasting rather than just one chapter. Econometric models are the most widely used of the analytical formulations discussed. They have long been established as forecasting tools for the airline and aircraft industries as well as other areas of the economy. In order to be compatible with the bulk of current models, and in order to be able to develop more advanced models in any area, it is necessary to have a firm grounding in econometric modelling.

A brief review of econometric theory and techniques will be presented. For a more complete picture, one should consult any of the numerous books on econometrics. In most cases, an equation or set of equations that are believed to in some way describe the level of an independent quantity are formulated as functions of exogenous variables and unknown coefficients. Regression analysis, which is essentially a curve fitting procedure, is performed in order to estimate the value of unknown coefficients.

The object of econometric analysis is to determine not just correlation, but also causation or at least explanation. Although this distinction might seem to be only semantic on the surface, in actuality, it is quite important, and can often be very critical. Correlation was discovered between lung cancer and heart attacks and cigarette smoking. The causal link was later established. When correlation was discovered between heart attacks and coffee drinking, many suspected that a causal link would be established. Further research showed that the correlation existed because heavy coffee drinkers were often heavy cigarette smokers.

Following is an example more closely related to air transportation. In a discussion with an advertising executive of a trunk carrier concerning his past advertising campaigns, a very successful advertising campaign was credited with significantly increasing the carrier's revenue in the early 1960's. It seems far more reasonable to expect that this increase resulted from a large decrease in fares in constant dollars and from a tremendous increase in level of service that was brought about by the introduction of jet aircraft. Although these examples might seem to be obvious after the fact, there are many subtle cases that could give trouble to even experienced forecasters.

For simplicity, most regressions are done on functions that are linear or can easily be linearized by a logarithmic or other simple transformation. Single equation systems include independent (exogenous) variables and dependent variables. Multi-equation systems include these as well as endogenous variables. For multi-equation systems, it is more convenient to think of the exogenous variables as predetermined variables. Systems dynamics which is discussed in another chapter, is a special case of multi-equation simulation.

One of the simplest and most common methods for estimation of the coefficients of a model is ordinary least squares (OLS). This procedure minimizes the sum of squared deviations of the prediction from the observed quantity. It is a minimum variance estimator, and under appropriate assumptions, also the maximum likelihood estimator.

The general form of a system is:

$$\mathbf{y} = \mathbf{x}\beta + \mathbf{e} \qquad (5.1)$$

$\mathbf{y}$ = vector of the observations of the dependent variable

$\mathbf{x}$ = matrix of observations of the exogenous variables

β = a vector of coefficients which are to be estimated
e = vector of errors from the model

To show that OLS is the best linear unbiased estimator, it is necessary to use the Gauss-Markov theorem and to assume that the y_i are independent, the error variance is constant, and the errors are independent with mean zero and are uncorrelated with y or x. Normality assumptions are often added.

It is desirable that an estimator be unbiased, efficient, and consistent. The expected value of an unbiased estimator is equal to the "true" value of the coefficient. An efficient estimator is one with a small variance, and hence a tight confidence interval. The estimated value of a consistent estimator will approach the true value of the coefficient as the sample size increases i.e. the bias and the variance both shrink to zero.

Error in the models comes from the stochastic nature of the real world, measurement error, misspecification of the model form, and/or non-inclusion of important variables. It is often assumed that relative to the y_i's, the X's are non-stochastic. However, one should justify this assumption before making use of it.

Many statistical problems arise from or cause biased or inefficient estimates of coefficients. If the error variance is not constant, heteroscedasticity exists. It is not an unusual situation, especially when dealing with cross-sectional data for larger errors to be associated with larger measurements e.g. a percent error rather than an absolute error. In dealing with time series data, it is a distinct possibility that measurements will become more accurate as time progresses. Bartlett's test and the Quandt-Goldfeld test can indicate when heteroscedasticity exists. Under conditions of heteroscedasticity, the estimated coefficients are not efficient.

When the error terms are correlated, serial correlation exists. The existence of serial correlation not only can bias estimates, but it can also bias statistical tests. Serial correlation is quite common when dealing with time series data. Cochrane-Orcutt, Hildreth-Lu, and Durbin-Watson procedures can help detect and correct these problems. The Durbin-Watson statistic (DW) is commonly used to test for serial correlation.

$$DW = \frac{\sum_{t=2}^{N} (\hat{e}_t - \hat{e}_{t-1})^2}{\sum_{t=1}^{N} e_t^2} \approx 2(1 - \rho) \qquad (5.2)$$

ρ = the first order correlation coefficient

A process that has no autocorrelation present might appear to be serially correlated due to measurement errors. Many real world measurement instruments experience drift, cyclical errors, or bias due to aging. If data is generated from such instruments, a false indication of serial correlation might be forthcoming.

When two or more variables are highly correlated, multi-collinearity exists. GNP and personal income are highly correlated, and therfore it is difficult to include both of them in the same model. As income in current dollars equal the price index times income in constant dollars, models that incorporate prices and income in current dollars tend to have collinearity problems. Multi-collinearity causes unstable and unreliable estimates. The estimated values of coefficients become very sensitive to sampling errors. Often computer accuracy and overflow/underflow problems cause routines to break down in the face of a high degree of collinearity. In extreme cases, the problem becomes one of estimating a system of n dimensions with knowledge of (n - 1) or less of the dimensions.

It is necessary to test the hypotheses implicit in models both for reasonableness and statistical validity. To test individual coefficients, the student's t-statistic should be used.

$$t = \frac{\hat{\beta} - \beta}{\sqrt{s^2 / \sum x_i^2}} \qquad (5.3)$$

s^2 = sample variance

x_i = variation from the sample mean of the individual variable to be tested

β = coefficient of variable to be tested

The most common test is to determine whether $\hat{\beta}$ is significantly different from zero. In such cases, β is set to zero, and t is compared against tabulated critical values of $t(t_c)$ for the desired confidence levels and the appropriate number of degrees of freedom. One sided tests can also be done i.e. is β greater than zero. The confidence interval at the desired confidence level is defined as:

$$\beta = \hat{\beta} \pm t_c \, s(\hat{\beta}) \qquad (5.4)$$

R-squared is a goodness of fit statistic.

$$R^2 = \text{explained variation / total variation} = 1 - \frac{\sum \hat{e}_i^2}{\sum (Y_i - \bar{Y})^2} \qquad (5.5)$$

R-squared is not a totally satisfactory variable for explaining goodness of fit. As the number of variables in a regression increases, the goodness of fit also has to increase if there is any correlation at all between the dependent variable and the new variables that is orthogonal to other variables. Even a nonsense variable could increase goodness of fit. The corrected R-squared ($\bar{R}^2$) is similar to R-squared, but takes account of the number of observations and the degrees of freedom.

$$\bar{R}^2 = 1 - \frac{\sum \hat{e}_i^2 / (n-k)}{\sum (Y_i - \bar{Y})^2 / (n-1)} \qquad (5.6)$$

n = number of observations

k = number of coefficients estimated

Even with the number of degrees of freedom objections removed, it is incorrect and very naive to compare two models based solely on corrected R-squared. It is better to accept a behaviorally valid model with a low $\bar{R}^2$ than to accept a less logical model with a higher $\bar{R}^2$.

The F-statistic is used to make tests similar to t tests, but on several rather than just one variable. As with the t-statistic, the F-statistic is a measure of statistical confidence.

For one variable, $$F = \frac{\hat{\beta}^2 \sum x_i^2}{s^2} = t^2 \qquad (5.7)$$

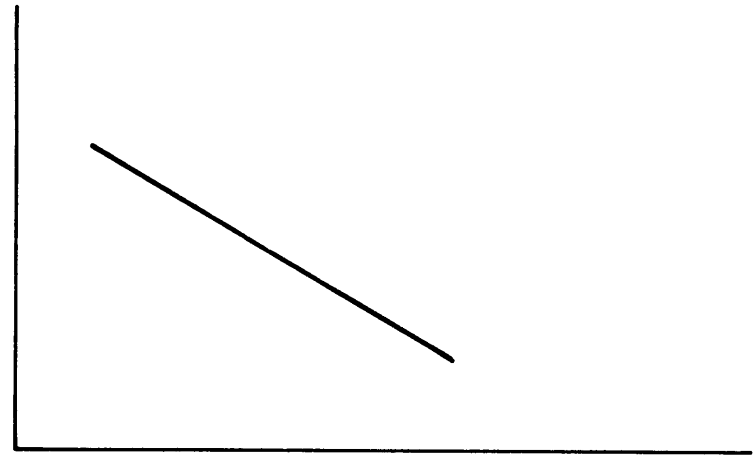
As with t, critical values of F are tabulated. Most regression packages print out values of F for the entire regression. This is almost useless as one would probably never create a model that did not have a high F-statistic for the entire set of variables. If there were two variables for which one had a priori reasons to believe belonged in a model, but neither of these variables had a significant t-statistic due to multi-collinearity, an F test could identify their influence. One variable could be dropped, or a new variable that is a combination of the two could be created.

It is sometimes useful to make use of dummy variables. These can help explain the effects of wars, strikes, seasonal variations, or other signficant departures from the norm. One should be wary of dummy variables that are added after the fact such as a regression of automobile sales that includes a dummy variable for the year that consumers got "tail-fin fever." A dummy variable for the effect of the opening of Disney World on air traffic to Florida was used in a model presented later in this chapter.

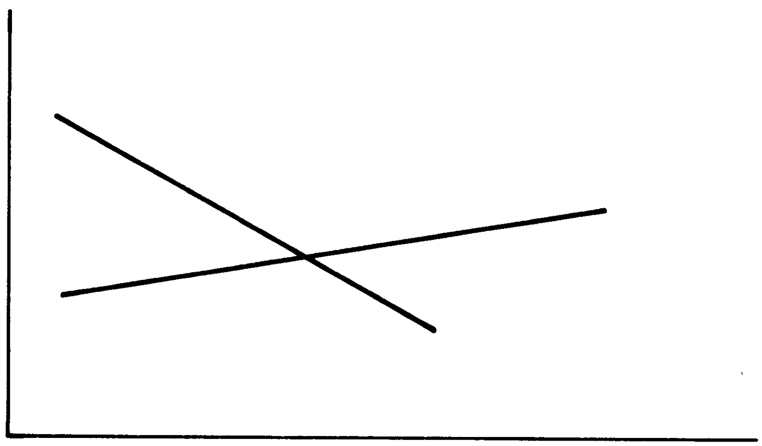
Quite often, statistical problems make ordinary least squares biased or inefficient. Other aggregate techniques exist which enable econometric models to be correctly estimated. Without going into details, these include the usage of regression of differences, addition of auto-correlated terms, indirect least squares (more of a pedagogic tool than a practical tool), constrained regression, two-stage least squares, three-stage least squares, weighted least squares, generalized differencing, and generalized least squares.

Identification problems occur in both complex and simple systems. As with any algebraic problem, in order to estimate n unknowns, one needs to have n independent equations. For instance, if one attempts to estimate a supply curve with a single equation, he will not be able to uniquely identify it because all that can be observed if a scatter of points that is some combination of the supply and demand curves. However, if demand varies considerably more than supply does, the curve that is estimated may appear to be the supply curve. In order to identify the system, one needs not only the supply equation, but also the demand equation. In general, the condition for identification of an equation is that the number of predetermined variables excluded from the equation, but in the system must equal the number of endogenous variables included on the right-hand side of the equation. Fewer excluded predetermined variables results in an under-identified equation. More excluded, predetermined variables than necessary results in an over-identified equation. Examples of over, under, and exactly identified equations are shown in Figure 5.1.

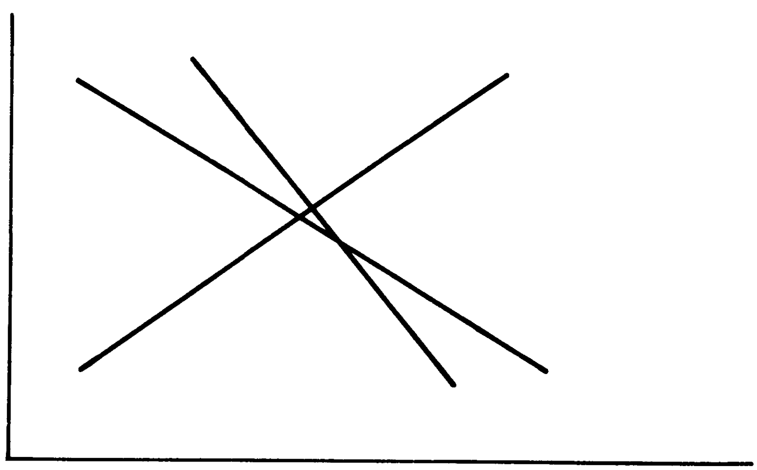
There are many linear and non-linear equation forms that could be used for aggregate demand models. Following are some of the more frequently used forms that are linear or easily convertible to linear. These functional forms will be examined



a) Under-identified (No intersection)



b) Exactly identified (Unique intersection)



c) Over-identified (Non-unique intersection)

Figure 5.1 a) Under, b) Exactly, and c) Over-Identified Equations with particular emphasis on the implied elasticities of demand.

$$Y = \sum \alpha_i X_i \tag{5.8}$$

$$E^Y_{X_i} = \alpha_i X_i / Y_i \tag{5.9}$$

This model form might be acceptable if it is hypothesized that the exogenous variables have independent rather than interactive effects. For positive coefficients, if X_i doubles, Y_i less than doubles, and $E^Y_{X_i}$ increases. There is no logical reason why this should be true for all X_i. Fare elasticity is expected to increase with increases in fare level; income elasticity increases and then levels off; other quantities might actually decrease. This model form is acceptable only for certain variables and over certain data ranges.

$$Y = \prod X_i^{\alpha_i} \tag{5.10}$$

$$E^Y_{X_i} = \alpha_i \tag{5.11}$$

This model is useful when certain interactive effects exist among the independent variables. The logarithmic transformation that is made in this and other models prior to solving could increase the problems of multi-collinearity or reduced accuracy on limited word length computers. The elasticities of this model are based on averages over the aggregate quantities involved, and these elasticities can change if the makeup of the market changes. There are usually not good reasons to assume constant elasticities over all ranges.

$$Y = \prod e^{(\beta_i X_i)} \tag{5.12}$$

$$E^Y_{X_i} = \beta_i X_i \tag{5.13}$$

For positive β_i's, $E^Y_{X_i}$ increases proportionally as X_i increases. As in the previous forms, this model and its elasticities are reasonable only for certain cases.

$$Y = \prod X_i^{\alpha_i} e^{(\beta_i X_i)} \tag{5.14}$$

$$E^Y_{X_i} = \alpha_i + \beta_i X_i \tag{5.15}$$

$$Y = \sum (\alpha_i X_i + \beta_i \ln(X_i)) \tag{5.16}$$

$$E^Y_{X_i} = (\alpha_i + \beta_i X_i)/Y \tag{5.17}$$

The above two models are combinations of previous forms. Although the forms of neither of these models makes sense for all variables over all ranges of data, their greater flexibility makes them potentially more suitable for many problems. However, the added complexity involved with these models can result in a greater likelihood of statistical or computer problems.

These represent some of the more frequently used linear (or transformable to linear) forms used for aggregate demand forecasting models. Other linearizable or non-linear forms might yield more reasonable elasticities and functional relationships. For many short and medium term forecasting models, these forms appear to perform well. It is probably better for many purposes to use disaggregate models or those with other functional forms that yield more reasonable elasticities or even impose no constraints on the elasticity of demand.

The gravity is a direct modal generation model that distributes traffic volume in a simultaneous structure as some function of impedance to travel (usually distance or time). It can be modified so as to be just a generation and distribution model.

Classical form: $$V_{ij} = \frac{K\, P_i^{\alpha_i}\, P_j^{\alpha_j}}{d_{ij}^{\,2}} \qquad (5.18)$$

General form: $$V_{ij} = \frac{K\left(\prod_{\ell} M_{\ell}^{\alpha_{\ell}}\right)\left(\prod_{n} LOS_n^{\beta_n}\right)}{I_{ij}^{\gamma}} \qquad (5.19)$$

K = constant

P_i = population of market region i

α, β, γ = parameters to be estimated

d_{ij} = distance between regions i & j

M_{ℓ} = mass variables that might consist of socio-economic and population variables

LOS_n = level of service variables not included in I_{ij}

I_{ij} = impedance between i & j

The classical form is similar in appearance to the Newtonian law of gravitation. Contrary to the beliefs of many, there is no valid reason (other than similar functional forms and variables) to believe that the gravity model is derived from Newton's law of gravitation (even ignoring relativistic considerations). Newton's law obeys the laws of physics in three dimensional conservative force fields for point masses. Population and travel volume are quantities in two dimensions (surface of the earth). There is absolutely no reason to believe that there is a conservative force field present. For city pairs such as New York - Washington, the radius of the market regions is large relative to the distance between them, and so they cannot be considered point masses. The gravity model can be solved by simple regression techniques. It is widely used, and often gives good results.

Logistics, Gompertz, or other S-shaped curves are often used to make long range forecasts. They require non-linear estimation techniques.

Gompertz: $V = \overline{V}b^{a^t}$ (5.20)

Logistic: $V = V_{max}/(1 + e^{-(a + b t)})$ (5.21)

V = passenger volume

$\overline{V}$ = average passenger volume

V_{max} = saturation passenger volume

t = time

a,b = coefficients to be estimated

Although the above examples are for passenger volume versus time, other variables can be substituted.

The logistics curve is one of the easiest S-shaped curves to visualize and to estimate. It might be easier to estimate V_{max} than $\overline{V}$, although this is not necessarily true. The most common usage of these curves is for making long term forecasts for quantities such as total domestic demand as discussed in Chapter II under qualitative or technological forecasting.

S-shaped curves could be used for studying the relationship between market share and frequency share. It is necessary to develop a variable that defines the value of a departure (V_d). For example, V_d could equal $(1 / (1 + n))$, or V_d could equal $(1 / 2^n)$; where n equals the number of intermediate stops. One might also wish to include competitive factors such as time of day or slots where no competitors exist. A more appropriate, but also more difficult measure to implement, is the ratio of non-stop to multi-stop times. The city pair observations can be segmented into categories based on the number of carriers in the market and the distance between the origin and destination. Enough categories should be formed that the market regions in each class can be considered homogeneous with respect to number of carriers and distance. Alternatively, the S-shaped curve can be modified by multiplication by or addition of other terms. However, one must then constrain the system so that the estimated sum of market shares for any hypotehtical market adds up to unity. In general, this is difficult to do, so one is better off with a functional form that can easily assure this. There exist many other uses for S-shaped curves. In many cases, these curves model our non-linear world better than linear or log-linear approximations do.

In recent years, disaggregate, stochastic models have come into much greater use. Multinomial logit is a disaggregate, stochastic model that is popular due to the underlying assumptions of its functional form and its relatively small data and computational requirements. It has been derived from a variety of different theoretical considerations including a choice axiom on a constant utility model and a specific distribution, random utility model. It should be pointed out that in this model, utility can only be defined up to an additive constant.

Following is the functional form of the logit model.

$$P(i:A_t) = \frac{e^{V_{it}}}{\sum_{K \in A_t} e^{V_{Kt}}} \qquad (5.22)$$

t = a behavioral unit e.g. individual, family, etc.

A_t = the set of relevant alternatives for t (for air transportation problems, these can usually be defined by frequency, mode, destination, route (non-stop, connections) and time of day, day of week, etc.)

$P(i:A_t)$ = probability that behavioral unit t will choose alternative i out of A_t

$V_{it} = V_i(\mathbf{x}_i, \mathbf{s}_t)$

$\mathbf{x}_i$ = characteristics of alternative i

$\mathbf{s}_t$ = socio-economic characteristics of behavioral unit t

Many estimation packages require a linear utility function. However, log-linear functions can be used if they are transformed outside of the routine.

$$V_{it} = \sum_{\forall k} (X_{itk}\, \theta_k) \qquad (5.23)$$

$\mathbf{x}_{it}$ = a vector of characteristics that are functions of $\mathbf{x}_i$ and $\mathbf{s}_t$

$\bar{\theta}$ = a vector of utility (or disutility) functions

$$\text{elasticity:}\quad E^{P(i:A_t)}_{X_{jtk}} = (\delta_{ij} - P(j:A_t))\, \theta_k X_{jtk}$$

where δ_{ij} is the Kronecker delta (5.24)

$$\delta_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

The problem can be formulated with any general form of utility function (including non-linear forms). However, it is very difficult to compute the derivative of the likelihood function for non-linear forms, and therefore these forms are not well-suited for inexpensive and quick estimation techniques. A routine was developed to find the parameters that maximize the likelihood function by using enumeration. This routine was designed with a special case in mind. Although the problem was extremely small, (two modes, three parameters, and twenty-four observations), this routine took considerably more computer time for estimating this tiny problem than the Cambridge Systematics routine which uses derivatives of the likelihood function took for estimating a problem with four modes, four parameters, and over four hundred observations.[12] The message from this is clear. The computational requirements of non-linear utility functions probably far outweigh the advantages of these structural forms.

Wilson derives the logit model by treating consumers as gas molecules and applying entropy maximization.[13] He states that the probability of a specific distribution of consumers occurring is proportional to the number of distinct states of the system that give rise to that distribution. The number of distinct states for various configurations is subject to appropriate constraints. Although this derivation is subject to several severe criticisms, it does result in the same functional form as the logit model.

Peat, Marwick, Mitchell and Co. derives the logit model from another point of view. This discussion will use their notation which is slightly different than that which was previously introduced. Their model is based on the following assumptions. "... (a) the modal split of each mode is between 0 and 1, and the sum of all modal shares equals unity; (b) modal splits are monotonic functions of the independent variables; and (c) if the transportation variables are expressed in units such that the disutility of travelling by a given mode is an increasing (decreasing) function of its transportation variables, then the shares of that mode decreases (increases) when any of its transportation variables increase (decrease)."[14] Item (c) is just a simple assumption on the elasticity of demand.

The assumptions lead to a set of partial differential equations of mode M(W_m) with respect to the i'th attribute of mode j(X_{ij}).

$$\frac{\partial W_m}{\partial X_{ij}} = \begin{cases} -\alpha_{ij} W_m W_j & m \neq j \\ \\ \alpha_{im} W_m (1-W_m) & m = j \end{cases} \tag{5.25}$$

The solution of this set of partial differential equations for all modes and all attributes leads to:

$$W_m = \frac{e^{\sum\limits_i (\alpha_{im} X_{im} + a_m)}}{\sum\limits_{\forall j} e^{\sum\limits_i (\alpha_{ij} X_{ij} + a_j)}} \tag{5.26}$$

a_j = mode specific constant

The explicit addition of mode specific constants can be eliminated by setting the appropriate elements of α to unity. For practical purposes, at least one of the a_j's needs to be equal to zero or else identification problems will occur during estimation.

For reasons unknown to the authors,this model does not explicitly contain socio-economic variables although they might be implicitly included. They can easily be added, and without otherwise altering the derivation, the final form will be identical to the form that was first presented in the discussion of the logit model. Peat, Marwick, Mitchell and Co. claim that this model can be estimated by either simultaneous least squares or maximum likelihood techniques. The latter is probably less subject to statistical problems andis also likely to be less time consuming. They further claim that for typical problems, approximately five hundred observations are necessary for good calibration. This number appears to be reasonable as the small models used to provide parameter values for the simulation program in the next chapter would not perform well with fewer than two hundred observations.

With a little bit of hand-waving, other models can be made to resemble logit. The following simple mode split model is used for pedagogical purposes in several courses taught by the M.I.T. Flight Transportation Laboratory.

$$MS_{ijk} = \frac{I_k T_{ijk}^{\beta_k} C_{ijk}^{\alpha_k}}{\sum_{\forall m} I_m T_{ijm}^{\beta_m} C_{ijm}^{\alpha_m}} \qquad (5.27)$$

MS_{ijk} = market share for city pair ij for mode k

I_m = image factor for mode m (a mode specific variable)

T_{ijm} = time between regions i and j on mode m

C_{ijm} = cost between regions i an j on mode m

β_m, α_m = time and cost elasticities respectively for mode m

This model is of the form:

$$MS_{ijk} = \frac{\prod_{\ell} B_{ijk}^{X_\ell}}{\sum_{\forall m} B_{ijm\ell}^{X_\ell}} \qquad (5.28)$$

If different behavioral units are accounted for, the ij subscripts are dropped, and aggregate market share is replaced by disaggregate probabilities, then this model becomes:

$$P(M:A_t) = \frac{\prod_{\ell} B_{Mt\ell}^{X_\ell}}{\sum_{m} B_{mt\ell}^{X_\ell}} \qquad (5.29)$$

$$\prod B_{mtl}^{X_\ell} = \exp(\ln(\prod_\ell B_{mtl}^{X_\ell})) = \exp(\sum_\ell (X_l \ln(B_{mtl}))) \tag{5.30}$$

Substituting (5.30) into (5.29) yields:

$$P(M:A_t) = \frac{e^{\sum_\ell (X_\ell \ln(B_{mt\ell}))}}{\sum_{\forall m} e^{\sum_\ell (X \ln(B_{mt\ell}))}} \tag{5.31}$$

This model has now become the logit model with the characteristics of alternative i for behavioral unit t replaced by its natural log. As the primary reason for using a linear utility function is based on computational requirements, this is not a significant departure from the multinomial logit model in its usual form.

Discriminate analysis provides another means of deriving the logit model. Consider several alternatives (A_i, i=1,...,n) and a set of explanatory variables (X). People choosing alternative A_i are distributed as in Figure 5.2. Assuming independence of alternatives:

$$P(A_i|X) = (A_i,X) \;/\; P(X) \tag{5.32}$$

$P(Y|Z)$ = conditional probability of Y given Z

$P(Y,Z)$ = joint probability of Y and Z

$$P(A_i,X) = P(A_i)\ P(X|A_i) \tag{5.33}$$

$$P(X) = \sum(P(A_i)\ P(X|A_i)) \tag{5.34}$$

substituting equations 5.33 and 5.34 into equation 5.32:

$$P(A_i|X) = \frac{P(A_i)\ P(X|A_i)}{\sum P(A_i) P(X|A_i)} \tag{5.35}$$

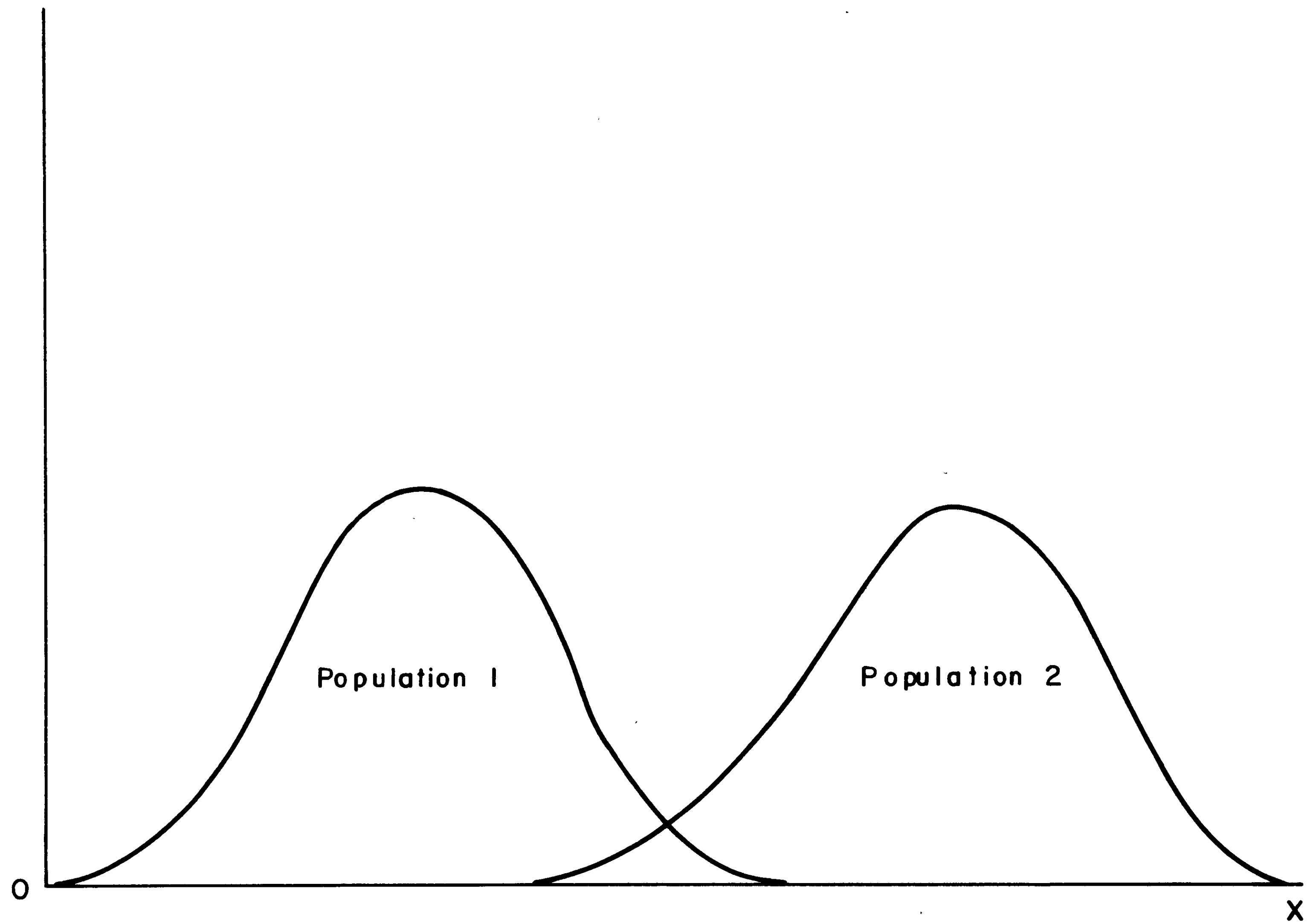


Figure 5.2 P. D. F. of Values of the Discriminant Function (Binary Case)

If normal distributions and a binary choice are assumed:

$$P(A_i|X) = 1 / (1+e^{-V+q}) = e^V / (e^V+e^q) \quad (5.36)$$

This is the same form as the binary logit model. V and q can be determined by estimating the means and the covariance matrices for the relevant modes and parameters.

Perhaps the simplest way of explaining the multinomial logit model is to examine the theoretical considerations of a binary choice model, and to then extend it to the n-dimensional case. One must first make the assumptions made in the PM&M derivation described previously. These assumptions are general and do not significantly restrict the model. For the time being, the subscript that defines the behavioral unit will be dropped.

$$P_1 = \alpha^T \mathbf{x} \quad (5.37)$$

$$P_2 = 1 - P_1 \quad (5.38)$$

$P_i (i=1,2)$ = the probability of choosing mode i

α = a vector of utility functions for each attribute

$\mathbf{x}$ = a vector of the difference between modes for each attribute

Observations of individuals will yield choice decisions rather than probabilities. One could run a linear regression using OLS or GLS on P_1 versus X, making sure not to include a constant term. (This simplifies the explanation; however, a constant term could be interpreted as a mode specific constant.) Non-zero error variance, as well as changes in X as time passes could cause P_i to be greater than one or less than zero. As probabilities only have meaning between zero and one, P_i must be restricted to those values. A rational individual will choose mode 1 if the difference in the utilities of mode 1 minus mode 2

$(V - \overline{\alpha}^T\overline{X})$ is greater than the random and unmeasureable effects of mode 2 minus mode 1.

Even if the estimated regression has the same slope as the unconstrained "true" regression, there would be heteroscedasticity problems caused by points outside of the zero/one range. This can easily be seen by examination of Figure 5.3.

To summarize this simple model:

$$P_1 = \begin{cases} 1 & V \geq V_1 \\ \dfrac{V - V_o}{V_1 - V_0} & V_0 < V < V_1 \\ 0 & V \leq V_1 \end{cases} \qquad (5.39)$$

One notices that there are discontinuities and singularities in the derivatives of P_1. To get around these problems, one needs to create a functional form similar to the one in Figure 5.4 (S-shaped curve bounded between zero and one). Probit models use an S-shaped curve that is created from a cumulative normal distribution that is shown in euqation 5.40.

$$P_1 = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{V} e^{-t^2/2} \, dt \qquad (5.40)$$

Logit models use the functional form shown in equation 5.41.

$$P_1 = 1 \,/\, (1 + e^{-V}) \qquad (5.41)$$

Logit derives its name from the similarity of its functional form in the binary case to the logistics curve. (If $(V_1 - V_2)$ is substituted for V, the following form results:

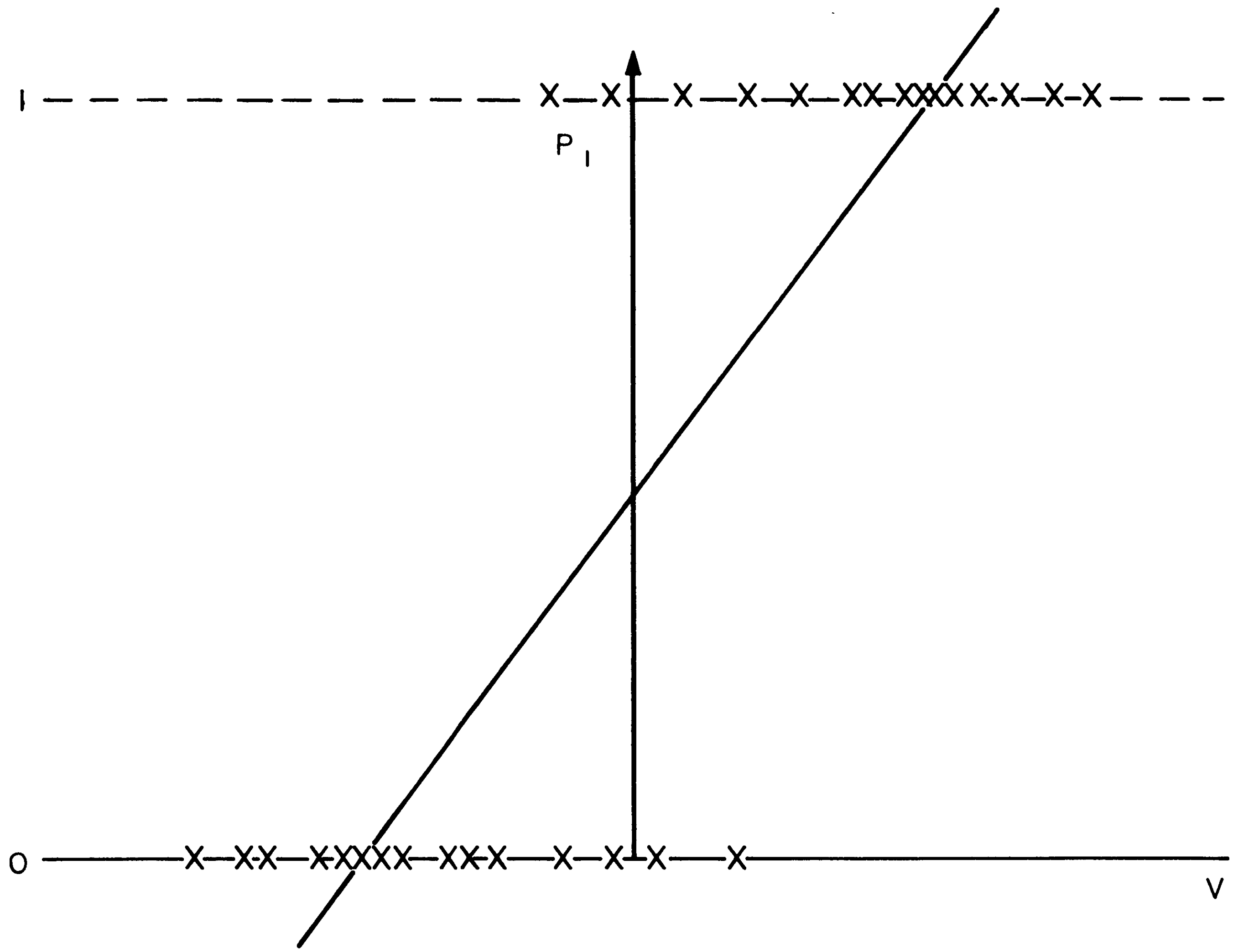


Figure 5.3 Linear Regression on Observations of Binary Choice Versus Utility (V)* $*V = V_1 - V_2$

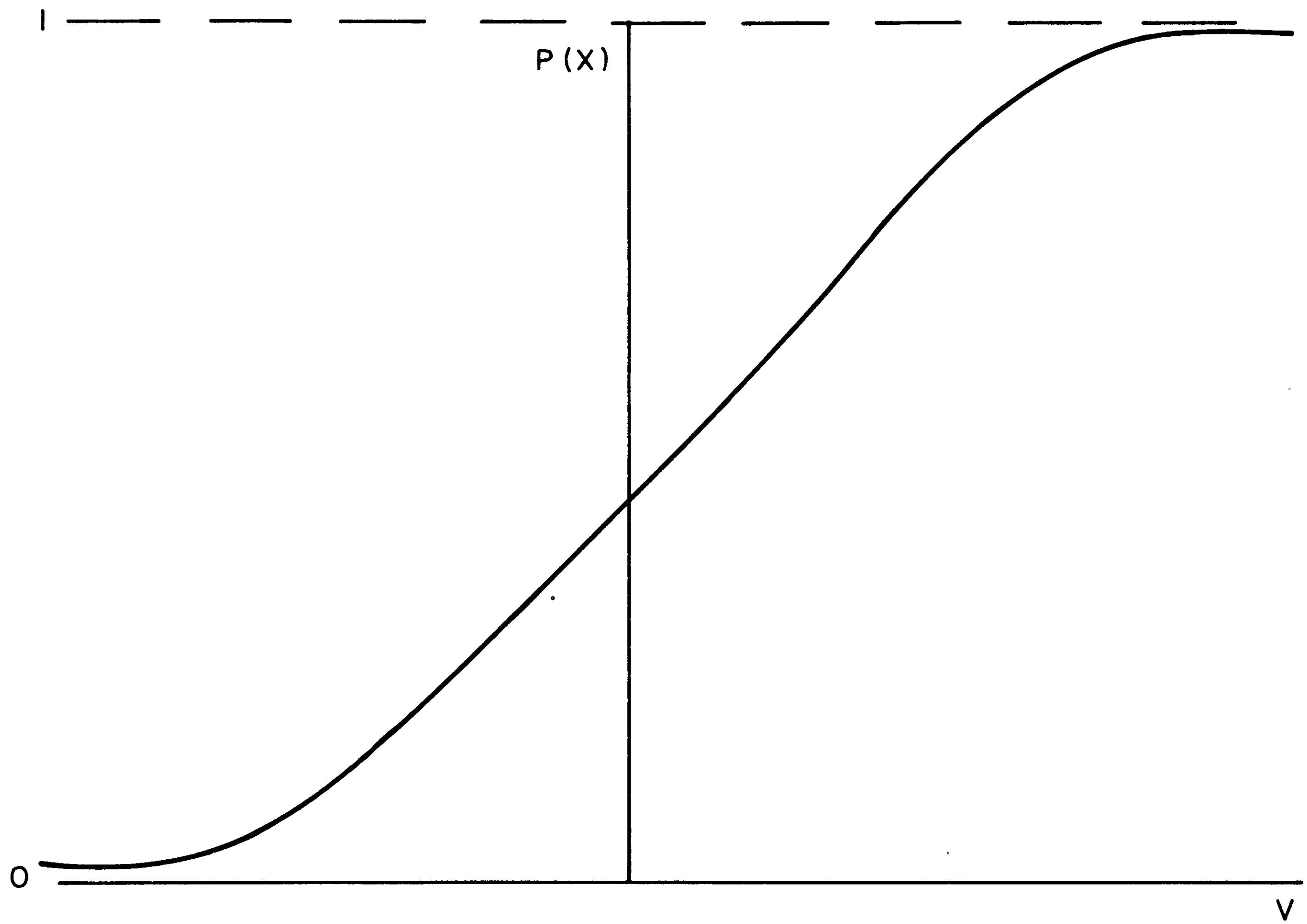


Figure 5.4 Functional Distribution for Disaggregate Choice Models

$$P_1 = \frac{e^{V_1}}{e^{V_1} + e^{V_2}} \tag{5.42}$$

From here, it is easy to see how the form of the multinomial case is arrived at.

$$P(i:A_t) = \frac{e^{V_{it}}}{\sum_{\forall k \in A_t} e^{V_{kt}}} \tag{5.43}$$

$$V_{it} = \sum_{\ell} X_{it\ell}\theta_\ell \tag{5.44}$$

However, for the normality assumptions to hold, rather than using the S-shaped form of the logistics function, it is necessary to use the following distribution: $e^{-ne^{-\nabla}}$.

Cross classification or category analysis is a simple technique that can prove useful for many classes of problems. A notable feature of this disaggregate technique is that it makes no assumptions on the functional form of the model or its elasticities. After deciding what variables influence air travel demand, one can then construct an n-dimensional matrix of these characteristics (e.g. income, family structure, occupation, location, etc.). N-dimensional volumes are formed by the intersection of specific ranges of these variables.

Through the use of a household survey, one can compute the number of trips per unit time that an average behavioral unit in each category makes. Each category has to be small enough that it can be considered relatively homogeneous. This is to insure that inter-category variances are large relative to inter-category variances. However, having too many categories results in extra work and less reliable estimates of the aggregate quantities.

One needs to adjust the volume of each category until a good compromise has been reached between homogeneity and extra work and unreliable estimates.

For making forecasts, one must predict the number of consumers in each category, multiply by the appropriate trip generating value, and sum the results. This procedure is in general not sensitive to changes in level of service. Rather than having a constant trip generation value in each block, one can derive relationships that are based on level of service. This can involve a lot of extra work for even small problems. Rather than doing this, one is probably better off using another disaggregate technique such as logit.

Many categories might be poorly represented or even nonexistent in the base year, but can be expected to be densely populated in the forecast years or in another market region. In such cases, this technique results in poor estimates.

Many of the usual statistical tests from econometrics cannot be applied to cross classification techniques, thereby causing them to be difficult to evaluate with respect to confidence and significance of estimates and of forecast quality. It could very well be necessary to collect more data for calibration of category analysis models than for other disaggregate models. Further hindering this type of model is the necessity for complete respecification and recalibration of the model should a new variable be added, or the structure otherwise changes. This technique, although relatively simple, has several important criticisms. For most purposes, a more advanced technique such as logit or probit is to be preferred.

In order to demonstrate the use of econometric models, aggregate models were created for the New York - Florida, New York - Orlando and total transatlantic travel markets. The first two direct demand models implicitly include trip generation, trip distribution, and modal split considerations. As they might be used for several purposes, it is desirable to create both long and short term models. These markets include some of the most densely travelled long haul domestic routes. Properly serving these routes is critical for several airlines.

The choice of variables is relatively simple for this case because many other models have been built for similar markets. GNP or average per capita disposable income, both in constant dollars, are candidates for inclusion. As the two are highly correlated, and since these are personal rather than business markets, it is most appropriate to include only income. Population obviously affects travel demand. One should attempt to define a population variable that accounts for the bulk of travellers in a properly weighted manner; in lieu of this, total domestic population was used. Most models include a price variable. It would be best to use total trip cost (including ground costs). For reasons of data availability, a measure of fare will be used. As air fare makes up a large percentage of total trip cost, this should be approximately correct. For simplicity, coach fare in constant dollars will be used. Other alternative price measures might include minimum fare or average yield. Seasonality will be accounted for in the long term models by a dummy variable called winter which is equal to one in the first quarter, and zero in all other periods. As the other periods do not depart much from each other, it is not really worthwhile to lose a degree of freedom to add other seasonal dummies. For multiplicative models, this dummy took on the values of e and one. The short term models took account of seasonality by using volume lagged one year. This variable also implicitly takes into account frequency and load factor, which are also believed to influence demand. It was also desirable to examine the effects of the opening of Disney World and of the weather on demand. The Disney variable equalled one when Disney World was open, and zero when it was closed. (Note: It is clear that Disney World affects travel to Orlando, but it is not clear whether it diverts traffic from other Florida destinations or if it generates entirely new traffic.) The weather variable was constructed by taking twice the number of sample standard deviations of the departure from normal of Miami temperature minus New York temperature plus the number of standard deviations of departure from the mean of New York precipitation minus Miami precipitation. Weather was assumed to have no effect in other periods and was therefore set to zero. It is by no means clear that this is the best form for the weather variable, but it does capture most of the effects. It later became necessary to set upper and lower bounds on the weather variable. As weather cannot be predicted more than one period ahead, it was necessary to limit the use of this variable to short term forecasts. For the multiplicative models, appropriate exponential transformations were performed on weather and Disney. Volume was computed from a ten percent sample of tickets taken by the Civil Aeronautics Board. The expected measurement error is 1 - 1.5% for the later Florida volumes and slightly greater for the Orlando volumes and the early Florida volumes. Florida volume was represented by summing the traffic at Orlando, Tampa, Miami, and Ft. Lauderdale. This by far makes up the bulk of Florida traffic.

The quarterly data was regressed from the first quarter of of 1962 to the present (second quarter of 1973). The starting date was picked to insure that the regression would be done entirely in the jet age (as opposed to the era of propeller-driven aircraft). Scatter plots were not made as their primary value is to graphically indicate correlation. As there is already strong a priori reasons to believe that the chosen independent variables influence traffic, it is not necessary to examine the scatter plots.

Following are comments based on the time series plots of the data. The Florida traffic volume is highly cyclical and has a distinct upward trend with a noticeable increase when Disney World opened. The Orlando traffic volume does not have the distinct winter peak of the Florida traffic. It has an upward trend and a tremendous increase when Disney World opened. The Miami coach fare was used in regressions on Florida traffic; and the Orlando coach fare was used in regressions on Orlando traffic. As the C.A.B. at various times used slightly different fare formulas, the fares move together, but are not exactly collinear. The fares in constant dollars increase several times, but for the most part, tend to decrease. Per capita income in constant dollars and population increase at roughly constant slopes that are identical within an additive and multiplicative constant to take account of the units. This high degree of collinearity forces population out of the models. One might argue that to take care of this problem, total disposable income (which is population multiplied by per capita income) can be used. This constrains the coefficient of population to be equal to the coefficient of per capita disposable income. Prior knowledge causes this hypothesis to be rejected. To accept it would bias the coefficient values. The weather variable appeared to be random and with no trend.

It was decided that four models would be necessary: long and short term models for both Florida and Orlando. The short term models need not include the seasonal variable. The long term models did not include the lagged demand or the weather variable. As a priori knowledge indicated that all of the variables tried should influence the traffic volume, it was decided to accept any variable that had a t-statistic significant at ninety-five percent and also had the "correct" sign.

The models were constrained to be either linear or multiplicative in form. Neither type is truly appropriate based on examination of the elasticities, but either should be suitable over the range of values in question. The multiplicative form was tried first as the variables were expected to have interactive effects rather than independent effects. Unfortunately, no multiplicative form gave satisfactory results, so it was necessary to try additive models. The models that were tried are presented in Appendix C. As the models are self-explanatory, only the final models will be discussed.

The algorithm used to find the best model for each purpose consisted of initially including all variables, and then systematically eliminating one variable at a time until an acceptable model was found. The variable with the lowest t-statistic was dropped unless strong a priori knowledge required its inclusion over another variable. This process is similar to step-wise regression, but works from all variables downward rather than one variable upward. The process continued until all coefficients were significant and all reasonable models were examined.

Short Term Florida Model

Volfla = b*Volfla(-4) + d*Disney + e*Farefla + f*Income (5.45)

	SER	CRSQ	b	d	e	f	DW	
1962 1- 1973 2	5.22E3	.9385	.6169	10476	-41390	1878	1.69	value
			.1077	3129	11421	450		σ
			5.73	3.35	-3.62	4.18		t
1966 1- 1973 2			.5255	10776	-64479	2593	1.68	value
			3.41	2.83	-2.12	2.74		t

The lagged dependent variable biases the Durbin-Watson statistic towards two, but it is still probably acceptable. As with all of the models presented, changing the regression period significantly changed the values of the coefficients. Breaking the model into several subperiods to test the stability of the coefficients, or performing other tests or making further changes is not worthwhile because this and all other models presented herein will give unstable results for reasons discussed later. As in the other models, there were no surprises as to the values of the included coefficients. It should be noted that weather could not be included at a significant level.

Long Term Florida Model

Volfla = d*Disney + e*Farefla + f*Income + h*winter (5.46)

	SER	CRSQ	d	e	f	h	DW	
1962 1-1973 2	4.98E3	.9441	19144	-94569	4172	10616	1.33	value
			2529	7050	191	1673		σ
			7.57	-13.4	21.8	6.34		t
1966 1-1973 2			15445	-1.44E5	5352	10679	1.80	value
			5.2	-7.53	11.6	4.88		t

In the long term Florida model, the seasonal effect and the Disney World effect have the greatest impact on the forecast, and the fare term the least. The Durbin-Watson statistic and the corrected R-squared are satisfactory, although the Durbin-Watson statistic is borderline. Once again, the coefficients are not stable.

Short and Long Term Orlando Model

Volorl = d*Disney + e*Fareorl + f*Income + h*winter (5.47)

	SER	CRSQ	d	e	f	h	DW	
1962 1- 1973 2	400	.9518	4172	-4498	182	-336	2.24	value
			192	727	16.9	134		σ
			21.7	-6.18	10.8	-2.51		t
1966 1- 1973 2			4139	-6284	220	-527	2.45	value
			17.9	-2.85	4.75	-2.71		t

This model is of exactly the same form as the long term Florida model. As lagged volume and weather are not significant, the long and short term models became one and the same. The seasonal variable turned out to be negative, but it is insignificant when compared to measurement error. As the time series plots suggested, Orlando traffic is not distinctly seasonal. The Durbin-Watson statistic and the corrected R-squared are satisfactory. The coefficients are not stable.

For all of these models, the F-statistic was highly significant. The standard error of the regression is small relative to the one sigma value of the error of the traffic volume measurement. Examination of time series and histogram plots of the errors reveal some bias, but not an unreasonable or unexpected amount.

The short term models are probably suitable for making one or two period forecasts. As fare and income will only change slightly over the short term, and since the structure is expected to remain roughly constant from one time period to the next, this model really does not reveal much new information. The Flight Transportation Laboratory time series model presented in Chapter III would probably give just as good results as these models except for a transient when Disney World first opened. Forecasts were made with these models, and the results are pretty much the trend growth with seasonal fluctuation that could have been eye-balled from a time series plot. These short term models would probably be of little use to airline planners. Their intuitive mental models could probably do just as well.

One cannot place a high degree of confidence in the long term models (except in a stable world), both with respect to the values of the structural coefficients and with respect to the models' abilities to make intermediate and long term forecasts. Although the models include the most important variables at highly significant levels, and the other statistical measures appear to be satisfactory, these models are clearly unsuitable for policy analysis. Their value is probably only slightly more than time series models.

The coefficients are not stable, and the additive form of the model does not seem totally reasonable. These models are policy sensitive at only the grossest levels. It seems reasonable to expect that population will affect traffic volumes. The existence of Disney World probably has a percent effect on volume rather than an absolute effect. It is generally accepted among airline, hotel, restaurant, and chamber of commerce officials that weather has a significant effect on tourism, and hence, on air travel in vacation markets. Multi-collinearity causes several important variables to be excluded from the models.

Numerous factors must be excluded from these and other aggregate econometric models. Frequency and load factor are known to affect volume. In some periods, seats were being rationed and availability and convenience of departure were low. In other periods, there has been an abundance of seats and departures. Strikes and other drastic events do not affect these aggregate models. Crashes and other disasters in at least the short term cause some consumers to shy away from a particular airline, aircraft, or even air travel. These models give no clue as to what would happen if a new Disney World type attraction opened, or what effect a package that includes Miami and Nassau would have on traffic. New aircraft have been continually introduced. These represent increases in comfort and slight increases in speed. During the first half of the 1960's, flying in jets had a status value. More recently, flying in wide-bodied aircraft had a status value.

It would be very useful to have a model that could predict the effects of flying a new aircraft (in particular, an SST), attracting major conventions to Florida, improving meals, adding carry-on luggage compartments, increasing advertising, moving a flight time, or numerous other marketing efforts. Aggregate econometric techniques have a great deal of difficulty in predicting unfamiliar situations.

Over the past twenty years, the structure of the market has changed considerably. Kennedy airport has become more important. Miami International has been modernized, but has also become more crowded. Ft. Lauderdale has been serving a great deal of the Miami traffic. Fewer non-stop flights have been scheduled from Boston in recent years; and these passengers have been routed through New York or other cities such as Atlanta or Washington. The characteristics of alternative destinations and modes have changed. Perhaps the most drastic is that some European fares have been reduced from very high levels to levels that are highly competitive with the Miami - New York coach fare. Not the least significant change in structure is the recent energy crisis. Patriotic feelings of conservation and the reduction of frequency will tend to decrease air traffic. However, the uncertainty of gasoline supplies will probably cause a significant number of automobile travellers to switch to air.

Evidence indicates that income elasticity is a non-linear function of income level. The same is true for fare. It is clearly incorrect to use a constant coefficient for these terms. Even if there existed only one fare level, as previously discussed, fare would not be an appropriate price variable. An average total trip cost would be inappropriate because of the wide variation in trip lengths and the many different trip purposes and itineraries. The problem is further compounded by the many different fare classes and quite often their various restrictions.

For these and many other reasons, an aggregate demand model is lacking in many areas. There is no claim being made that the models developed herein are the best aggregate models that could be created. However, it is difficult to imagine that even the best possible aggregate demand model can be very policy sensitive or even have unbiased structural coefficients (in the real world, not statistical sense) relative to disaggregate models. The aggregate models that were created for this exercise have their value, but at the same time, are severely crippled by aggregation problems. During unstable times, they can best be used as supplements to judgemental models.

Perhaps the criticism of these models has been too harsh. They appear to be satisfactory for predictive purposes as long as the user is aware of the limitations forced by the aggregate form of the models. It would be far better to create a disaggregate model to predict Florida travel. Not only are the behavioral assumptions more reasonable, but the disaggregate form will eliminate many of the problems of multi-collinearity. On the negative side, much of the data that is necessary to calibrate a disaggregate model is proprietary or does not exist. However, for models of this size, the data collection costs will be small relative to the potential advantages.

As a further example of the use of econometric model to forecast the volume of passengers on a given route, we will use the case of the total North Atlantic market. The simplest and the most practical predictor variables to include in the aggregate model for the North Atlantic are GNP and average income per capita [15] to represent the socio-economic factors and average fare and trip time for the transport variables. In addition, it is normal to include a time-trend term to account for all the forces which should be explicitly included in the behavioral demand model but are unquantifiable for subjective.

Air travel demand is strongly determined by income--personal income in the case of pleasure travel and GNP in the case of business travel. The higher the income level of an individual, the greater is the likelihood of that individual traveling abroad for non-business reasons.[16] In 1963, 42 percent of the traveling U.S. citizens and 25 percent of the traveling Europeans had incomes above $15,000. By 1966, these percentages of the traveling population had increased to 60 percent for the U.S. and 31 percent for Europe. Various studies[17] have shown that a factor which is even more important than the level of personal income is the distribution of family income. The Port Authority of New York and New Jersey[18] survey data show that, whereas in 1956 seventeen percent of the American traveling population had income less than $5,000 per annum, in 1966 only 7 percent of this population segment were earning below that amount. The story is equally impressive for the European travelers: the percentage of the population with less than $5,000 income per annum fell from 33 percent to 19 percent in the period from 1963 to 1966. Furthermore, while the money at the disposal of Europeans, even in the most prosperous countries, is still considerably below that of U.S. residents, the gap is tending to narrow, implying that the ratio of European air passengers to American air passengers on the North Atlantic may change in the future.

Some analysts prefer to use the distribution of family income above a certain base level. Such a distribution, although logical, is difficult to justify for three reasons. First, the base level is a subjective measure and analysts differ in their views of its numberical value. Moreover, in the case of North Atlantic travel, the level would vary by country. Second, the data are very fragmentary on the distribution of income, especially for some of the European countries. Third, the variation in income distribution is fairly difficult to forecast accurately. Nevertheless, some analysts have performed extensive research in this area. The National Planning Association study considered such measures of inequality for comparison among income size distribution of different groups and different time period as Pareto's coefficient of inequality and Gini's concentration ratio based on the Lorenz curve.[19]

Business travel appears to depend, among other things, on GNP and particularly on exports, imports, the level of investment abroad and the balance of payments. It stands to reason that during recessions the amount of business travel diminishes. Conversely, during an expansion of the economy business travel increases. It can be seen from this that a relationship exists between the fluctuations in the economy on both sides of the Atlantic and the traffic trend. The National Planning Association study considers the ratio of pre-tax corporate profits to full-employment gross private product as an index of annual fluctuations in the economy. However, in simple analysis, although traffic is influenced by many factors simultaneously, it is common to indicate the relationship between traffic and the economy through the use of GNP as the predictor variable. The advantages of using GNP are as follows:

1. The economic conditions of two countries can be compared

on a consistent basis.

2. Fairly accurate historical data are available.

3. Many learned individuals and institutions have produced long-range forecasts of GNP to a high degree of accuracy using the most sophisticate methods available. Furthermore, these long-range forecasts are continuously reviewed and updated.

It is a common procedure to incorporate a time lag to allow for the elapsed time between the movement in the economy and its influence on traffic. For simplicity, this time lag is fixed. However, sophistication can be introduced through the use of a distributed lag technique whereby the influence of a change in the predictor variable may be felt over a longer period and in different amounts during each successive period.

Both personal and business demand for air travel are dependent upon the total trip cost and vary inversely with trip cost. Apart from slight fluctuations, the transportation cost for a North Atlantic trip has been reduced from about $600 in 1951 to about $450 in 1968. This cost represents the average fare for sea and air travel.[20] The total cost of a transatlantic trip has been decreasing due to the reduction in fares and the decline in average expenditures while traveling in Europe. The downward trend in expenditures abroad is explained partially by the growing number of U.S. citizens with limited funds who are now traveling and partially by the fact that air travelers have been staying shorter periods in Europe and spending less. The average stay has declined from about 66 days in 1950 to 45 days in 1963 and was estimated at 28 days in 1969.[21] Table 5.1 compares the major components of the cost of a ten-day trip in Europe and in a large city in the United States for the years 1958 and 1970. In both cases, the air fare represents a smaller part of the total cost in 1970. Although it is simpler to use the fare as a cost term in the model,

Table 5.1

Components of Cost of Travel

Component	Distribution of Expenses for a 10-Day Trip			
	In Europe		In a Large U.S. City	
	1958	1970	1958	1970
Air Fare	75.8%	48.7%	31.6%	18.7%
Meals	12.0	25.3	26.2	32.2
Hotels	12.2	26.0	42.2	49.1
Total	100.0%	100.0%	100.0%	100.0%

Source: Air Transport 1971. Air Transport Association of America, Washington, D.C.

it is more accurate to use the total cost which includes ground expenses in addition to the fare. When interpreting the price elasticity of demand, the analyst should differentiate between the two cases.

With the multitude of fares available on the North Atlantic at any given time, the selection of a particular fare becomes a complex problem. A typical fare structure would consist of a breakdown by season, shoulder, peak and basic in addition to the many excursion fares available. The excursion fares presuppose a given length of stay and are limited to certain times of the year. In theory, the model should consider the trend in the lowest fare, since this is the fare which affects the total size of the market, while changes in other fares affect the traffic mix. However, since there are a number of restrictions placed on the lowest fares, it is not feasible to use them. The analyst is, therefore, forced to use an average fare, even though very few passengers actually pay the average fare. The use of an average fare based on an average passenger yield (revenue per revenue passenger-mile) and a specified stage length presents some problems. For one, yield figures are only available for the scheduled carriers. Moreover, care must be used in interpreting the results of studies based on such fare data.[22] Strictly speaking, the regression coefficient obtained by using the average fare will not be the true price elasticity of demand and, as such, may not be appropriate for use in determining pricing policy or in planning market strategy.

The average fare as a representation of the price index is simplified even further by analyzing the trend of the average fare on a specific route such as New York to London. The assumption that the New York - London fare trend is a reasonable representation for the whole North Atlantic can be justified on the basis that almost all fares are "pegged" to this route, as has been historically true in the case of IATA members operating on the

North Atlantic. When a new fare is introduced on a particular route, such as New York - Rome, then the New York - London fare and almost all other fares are changed accordingly. Finally, modal competition from sea transport is normally not incorporated in the calculation of average fare. The justification offered for using only the average air fare is that air fare and sea fare are somewhat interrelated.[23] Also, sea travel is no longer considered a substitute for air travel. The passengers who now travel by sea do so for reasons other than cost. It may be for the sheer pleasure and relaxation of spending five days at sea or it may be from fear of flying. In either case, the modal choice is not dependent on the cost of the trip.

The downward trend in air fares has been important in attracting new travelers and in causing experienced travelers to take air trips more frequently. In addition, charter travel has played a very important role in the development of air travel, especially in the international market. Charter sales have increased with the corresponding increase in the price spread between charter service and scheduled service. The market share of the charter carriers on the North Atlantic increased from 16 percent in 1963 to 30 percent in 1971.[24]

The total demand for air travel (pleasure and business) varies inversely with the time required to complete a given trip. Reduction in trip time, basically due to the higher speeds of aircraft, has affected both the business traveler and the pleasure traveler. Higher speeds have meant that the businessman can reach his destination in less time. Higher speeds also mean that the pleasure traveler can visit more distant places in a given time. The parameter "travel time" typically consists of line-haul travel time plus some combination of access, egress, terminal processing and schedule delay time. The simplest way to incorporate the trip time in the model is to consider the average aircraft speed on the North

Atlantic at any given time. Since the average aircraft speed varies by market at any given time, it is necessary to weight each flight by the corresponding aircraft speed.

The choice of the general form of the demand model will depend primarily on such factors as the historical traffic trends, data considerations, the time period of the forecast and certain desired properties of the demand function, such as a constant or a variable price elasticity of demand.

The aggregate model assumes that the service--air travel--is an homogeneous unit measured in passengers, revenue passenger-miles or revenue ton-miles and that the volume of passenger traffic is related to the same variables in all markets. This implies that the travel demand in the New York - Lisbon market can be characterized by the same variables as in the New York - London market. The aggregate demand model does not stratify traffic by mode, class of service or purpose of trip. Additionally, the aggregate demand model generally does not contain a supply parameter. This is justified on the grounds that the airlines usually operate with considerably less than full capacity and it is therefore unnecessary to include a supply variable. Furthermore, monopolistic routes on the North Atlantic are almost nonexistent, and insufficient capacity is unlikely, due to the market forces. The standard criticisms evoked by excluding the supply factor are, first, that there may be some routes with very high load factors and, second, that an increase in supply may increase demand.

As an illustration, we will present the results of an aggregate demand model, calibrated using multiple regression analysis. Before evaluating the empirical results, it is necessary to be aware of three fundamental assumptions underlying this approach. First, it is assumed that most of the variation in the dependent variable can be explained by using a few selected independent variables. This assumption is necessary due to the fact that we have limited data. Besides, in many cases it is difficult, if not impossible, to quantify all the variables--even though we know that these variables have influenced the travel demand in the past and will continue to do so in the future. The second assumption is that it is easier and/or more accurate to forecast the independent variables. This is a very critical assumption, since the forecast of traffic cannot be superior to the forecast of the independent variables. The third assumption is that the functional relationship will remain valid throughout the forecast period.

The volume of air travel varies a great deal among the different city-pair markets on the North Atlantic. On-line traffic estimates in 1969 ranged from about 80,000 passengers a month in the New York - London market to about 600 in the Philadelphia - Frankfurt market.[25] It is clear that the population size of the origin and destination cities,the level of family income, the community of interest and the level of fare are the major determinants of the volume of traffic in a given city-pair market. A forecast of the individual city-pair or airport origin and destination traffic can generally be obtained through the use of a gravity-type model with a slightly different input as compared to the aggregate demand model described earlier. In the basic gravity model, the passenger demand between the two cities is hypothesized to be directly proportional to the product of the populations and inversely proportional to the distance between them. The constant of proportionality represents the community of interest between the two cities. Since the distance term represents impedance to travel, it is a common procedure to substitute trip time and/or fare in place of the distance factor. In this case, the ideal input would be to include total trip time, which consists of airport-to-airport trip time plus access and egress time plus schedule delay time, the latter being a function of the frequency of service. Again, the limitations of statistical data force the analyst to use airport-to-airport time only.

The indirect impact of frequency of service has already been mentioned in referring to schedule delay time. However, the demand for air travel on a specific city-pair is a direct function of the level of service offered. The level of service can be estimated by determining the weighted sum of frequencies for the city-pair. This procedure would be suitable for a new route. For an established route, since frequency is a function of demand, the simplest way to incorporate frequency is to include an auto-regressive component, namely, the demand in the previous time period, as one of the independent variables.

A significant parameter, which has been overlooked in many of the past forecasts, is the influence of the airline route pattern on the flow of air passengers on a segment. Since major airports serve large metropolitan regions rather than just the surrounding cities, individual city-pair demand and total city volumes are a function of the total airline service offered at a given time. For example, a high percentage of the on-line traffic between New York and London is connecting traffic. This traffic flow is, therefore, a function of the availability of direct service from other cities. For individual city-pair models, an attempt should be made to incorporate the level of service, not only on the route under investigation, but also on alternative routes. In this respect, it is also necessary to take account of the attractiveness of alternative destinations.

Although our example does not include it, the specification of the model should be dynamic in that it should reflect fully the time difference between a change in the independent variable and the accompanying response in the dependent variable. The application of this property is desired in differentiating between the short-term and long-term effects on demand response when changes are made in the fare. This can be achieved by using a technique developed by Marc Nerlove[26] for estimating long-term elasticities from time-series data, based on assumptions with regard to the "elasticity of adjustment" of quantity demanded to changes in price. The application of this technique is illustrated in the study performed by Watkins and Kaylor[27] on scheduled international air traffic of U.S. Flag carriers. Their results show that the application of Nerlove's technique produces long-run price and income elasticities of -1.5 and +1.9 respectively, compared to their short-run values of -1.044 and +1.293.

The format selection of independent variables is largely dependent upon the availability of historical and projected values for the data. For example, there are at least three forms of income that can be entered into the demand equation: national income, disposable income and discretionary income. Although the latter would be a more logical predictor variable of air travel, its use is constrained due to its subjective value, the difficulty of its quantification and the unavailability of consistent data, especially for many of the European countries. On the other hand, the United Nations annually publishes data on national income in consistent form for the European countries as well as for the United States. The question of format also entails a decision as to whether incomes should be expressed in current or constant dollars.

In the case of North Atlantic travel, it is desirable to weight the values of income and GNP to show the relative traffic-generating capabilities of origin and destination cities. One such system would be to weight each European country's income by the percentage of total transatlantic passengers generated by that country in a given year. The following formulation is an example of an aggregate demand model to forecast the long-term traffic on the North Atlantic route.[28] Although some sort of weighting system appears to be a superior formulation over a straight measure of income and GNP, the analyst should be aware of the need to forecast a similar weighting system for predictive purposes.

$$T_{ij}(t) = A \cdot G^{\alpha}(t) \cdot D^{\beta}(t) \cdot F_{ij}^{\gamma}(t) \cdot V_c^{\delta}(t) \cdot (1+g)^t \qquad (5.48)$$

where:

$T_{ij}(t)$ = total passenger traffic between i (U.S.) and j (Europe) during period t.

$G(t)$ = composite GNP in period t. This is determined by using the same weighting system as for composite income per capita described below.

$D(t)$ = composite national income per capita in period t.

$F_{ij}(t)$ = average air fare for traveling between i and j.

$V_c(t)$ = average cruise speed of aircraft in operation on the North Atlantic at time t.

$(1+g)^t$ = function of time trend.....a natural growth term. This implies that if GNP, income, fares and speed of aircraft were constant, the traffic would grow at g percent due to all other factors such as population, improvement in service and effect of variation in tastes.

$t = 1, 2, 3 \ldots\ldots$

$D(t) = D_{u.s.}(t) \cdot i_{u.s.}(t) + D_e(t) \cdot i_e(t)$

Where $i_{u.s.}$ and i_e are the percentages of total transatlantic traffic (European and U.S. residents) accounted for by U.S. residents and Europeans respectively in the year t.

$D_{u.s.}$ = income per capita of United States.

$D_e(t)$ = income per capita of Europe

$= D_1(t) \cdot i_1(t) + \ldots\ldots\ldots + D_n(t) \cdot i_n(t)$

Where $i_n(t)$ is the percentage of round trip transatlantic European traffic accounted for by the nationals of country n in the year t.

The major sources of statistical data on traffic volumes are the international trade agencies, government agencies and civil air transport agencies. A valuable source for aggregate traffic data on the North Atlantic is the International Air Transport Association's "World Air Transport Statistics." On-line traffic data for international city-pairs can be obtained from the "Traffic Flow Statistics" published by the International Civil Aviation Organization. Other data sources include agencies such as the U.S. Immigration and Naturalization Service, the U.S. Civil Aeronautics Board, the U.K. Civil Aviation Authority and the European Civil Aviation Conference.

Major sources of socio-economic data for the European countries are the reports published by the Organization for Economic Cooperation and Development and the reports published by the Bureau for Program and Policy Coordination of the Agency for International Development. The report, "Gross National Product Growth Rates and Trend Data by Region and Country," is of particular interest here. In addition, various issues of the United Nations Demographic Books can be very useful. Finally, the United States Department of Commerce is an excellent source of relevant U.S. data.

The calibration stage involves the empirical manipulation of various functional relationships for a base period. The objective is to find the relationship which gives least variance between the derived demand and the actual demand. The calibration of the model, that is, the estimation of the demand coefficients, is normally performed by employing multiple regression analysis. The coefficients in a standard single-equation, such as the one discussed above, can be estimated using ordinary least-squares. Multi-equation or simultaneous-equation models, on the other hand, require a more sophisticated calibration process. In this case, the coefficients can be estimated using techniques such as the reduced form, two-stage least-squares and maximum likelihood.[29]

Sometimes calibration is performed through conditional regression analysis. This refers to the calibration process wherein one of the independent variables in the equation is given a fixed value while the remaining coefficients are derived through the normal process. Conditional regression is normally used when multi-collinearity is a serious problem or when prior knowledge indicates a particular value of a given parameter. The existence of intercorrelation between price and income produces biased results for both of their respective elasticities. Watkins and Kaylor applied the conditional regression technique to the U.S. international market with an income elasticity fixed at +1.292.[30] This produced a statistically significant value for price elasticity of -1.257.

The model shown in Equation 5.48 was calibrated using historical data from 1951 through 1969. The results of the regression analysis were found to be unreliable and statistically insifnificant. The standard errors of the regression coefficients were relatively high. The coefficient of the GNP term carried a negative sign. The most significant term in the equation appeared to be the time trend. This is quite common, and many analysts rely on this dominance and forecast using time trend only.

The basic problem with the model was found to be the existence of high correlation between GNP, income, fare and time. The problem was partially eliminated by determining the statistical insignificance of the GNP term through the use of the Chow Test.[31] It is quite reasonable to discard the GNP term and let income explain a significant part of the variation. This can be justified on the grounds that almost 75 percent of the North Atlantic travel is for pleasure and personal reasons for which income appears to be a more reasonable explanatory variable. The time trend was also eliminated, since it biased both the income and price elasticity of demand. Finally, although the impact of aircraft speed was very strong with the introduction of jets, its influence stabilized towards the second half of the decade; therefore, the term was left out of the model.

The model shown in Equation 5.48 was reformulated in accordance with the preceding discussion. Table 5.2 shows the empirical results. Both the income and the fare terms are statistically significant and logically acceptable. Judging from an application of the usual statistical tests, the model appears to be valid as a forecasting tool. However, price and income elasticities (although statistically significant) are not necessarily unbiased due to the intercorrelation between these two parameters. This is suspected because of the high R-squared value and the relatively high standard error of the coefficients. While it may be invalid to compare the regression coefficients from different models, it is worthwhile to note that the National Planning Association study[32] surveyed the literature extensively and estimated the income elasticity to be +1.4 for U.S. residents and +1.2 for European residents, while the price elasticity was estimated at -1.2.

As mentioned previously, these elasticities represent average values over the range of data from 1951 through 1969. They are very sensitive to the specification of the model and the time period

Table 5.2

Regression Coefficients

Independent Variable	Estimated Coefficient	Standard Error	t-Statistic
Ln K (constant)	3.529	4.786	0.737
β (Ln Income)	1.555	0.416	3.740
γ (Ln Fare)	-1.182	0.280	-4.222

R-Squared = 0.988

F-Statistic = 671.032

Durbin-Watson Statistic = 1.811

Number of Observations = 19

Sum of Squared Residuals = 0.074

covered. In order to judge the stability of these coefficients, the same model was calibrated twice, using data from 1951 through 1959 and from 1960 through 1969. For the former years, the price elasticity was observed at -2.670; for the latter time period, its value was -0.388. The income elasticity took on the reverse trend. It was insignificant in the former case at +0.357 and highly significant in the latter case at +2.507. In the former time period, the price variable tended to explain a significant part of the influence of the income factor. The reverse was true for the latter period. This illustrates the limitation of the average values of the elasticities and the lack of confidence which can be placed on the validity of the demand model as a tool for market strategy planning or determining the impact of alternative pricing policies.

Chapter VI

Simulation Models

For the most part, the use of simulation models for air travel forecasting has been minimal. However, the use of simulation models in other related and unrelated fields is rapidly expanding; and undoubtedly, this trend will soon be affecting air travel demand models. A simulation model is one that makes use of known or hypothesized relationships to reproduce the actions of a system through time or some other dimension. Simulation models can make use of virtually any analytical or dynamic relationships. One of their primary advantages is that closed form solutions (i.e. exact mathematical solutions) are not always necessary for their use. Closed form solutions can be approximated by stepwise calculations.

Models can be either discrete or analog, although discrete formulations are usually preferred for computational reasons. Discrete formulations deal with distinct and quantized (digital) quantities. The fineness of division should be such that for all practical purposes, limitations forced by discreteness can be ignored. Analog formulations deal with continuous quantities. The capability to do so usually is not required with the availability of high-speed, digital computers. Analog simulations tend to suffer much more than discrete simulations from noise in the system.

The increase in the use of simulation models is due directly to the availability of quick and inexpensive computer facilities. An eight-day space mission complete with guidance, navigation and control systems can be simulated overnight. Prior to the age of computers, such a simulation would have taken many man-years.

In designing a simulation model, one should identify all potentially important variables and relationships. The model should be structured into various subprograms. Each subprogram should serve just one purpose. It will be easiest to understand the system if logic sections and equation sections are not mixed in the same module. The program will fit together much better if each module has only one entrance and one exit. It is not necessary to understand the actions of the entire system if one can understand the relationships among all inter-connected modules. Presumably the model itself will enable the user to gain a greater understanding of the system.

With a properly structured program, additional variables and relationships usually can be added without requiring major revisions. This feature makes the testing of alternative plans or alternative scenarios much easier. It also permits a high degree of flexibility with respect to accuracy, detail, and aggregation.

The problem is then solved in a step-by-step manner. The simulation can proceed as a function of time or can be designed to perform other non-dynamic activities. Simulations can be programmed in specially designed languages such as GPSS, SIMSCRIPT, or DYNAMO. These languages have built-in features that make programming simulation models easier. This and other features enable a person who has only a bare minimum of computer experience to create simulation models. It is not necessary to limit oneself to special purpose languages. In fact, it is often better for reasons of flexibility and program efficiency to program a simulation model in a general purpose language such as Fortran or PL-1.

One usually needs to define discrete or continuous probability density functions in order to account for the stochastic properties of many relationships. Probability density functions (PDF) specify the likelihood of an event occurring. Probably the best way to explain them is by an example. Figure 6.1 shows a continuous probability density function. The probability that a random variable will have a value between a and b is simple the area under the curve between a and b in mathematical terms:

$$P(a \leq X \leq b) = \int_a^b f(x)\,dX \qquad (6.1)$$

A discrete probability density function is shown in Figure 6.2. The total area under under a continuous probability density function or contained in the bars of a discrete probability density function is always equal to unity. When the stochastic parameters are generated through independent, probabilistic methods, this technique is called Monte Carlo simulation because of the chance nature of the process.

The following example of Monte Carlo generation of characteristics should nelp clarify the technique. This example can be part of a model of an air transportation network. The network can be broken into links. Several stochastic events determine the time for the aircraft to traverse the link. The probabilistic events can be modelled in several ways but usually they are characterized as a random variable with some specified mean and standard deviation. The time that the door is closed and the blocks are removed is a function of when the aircraft and crew arrived, passenger processing time from gate to aircraft, and time to perform maintenance such as refueling, cleaning up, and stocking the galley. Once these times are generated, one can generate the time between block removal and start of take-off. This is a function of time of control tower approval, distance to the taxiways and then the runway, number of aircraft ahead in line, and ATC rules (e.g. landing aircraft are given precedence over departing aircraft in dual purpose runway configurations if conflicts exist). Each element of this time can be

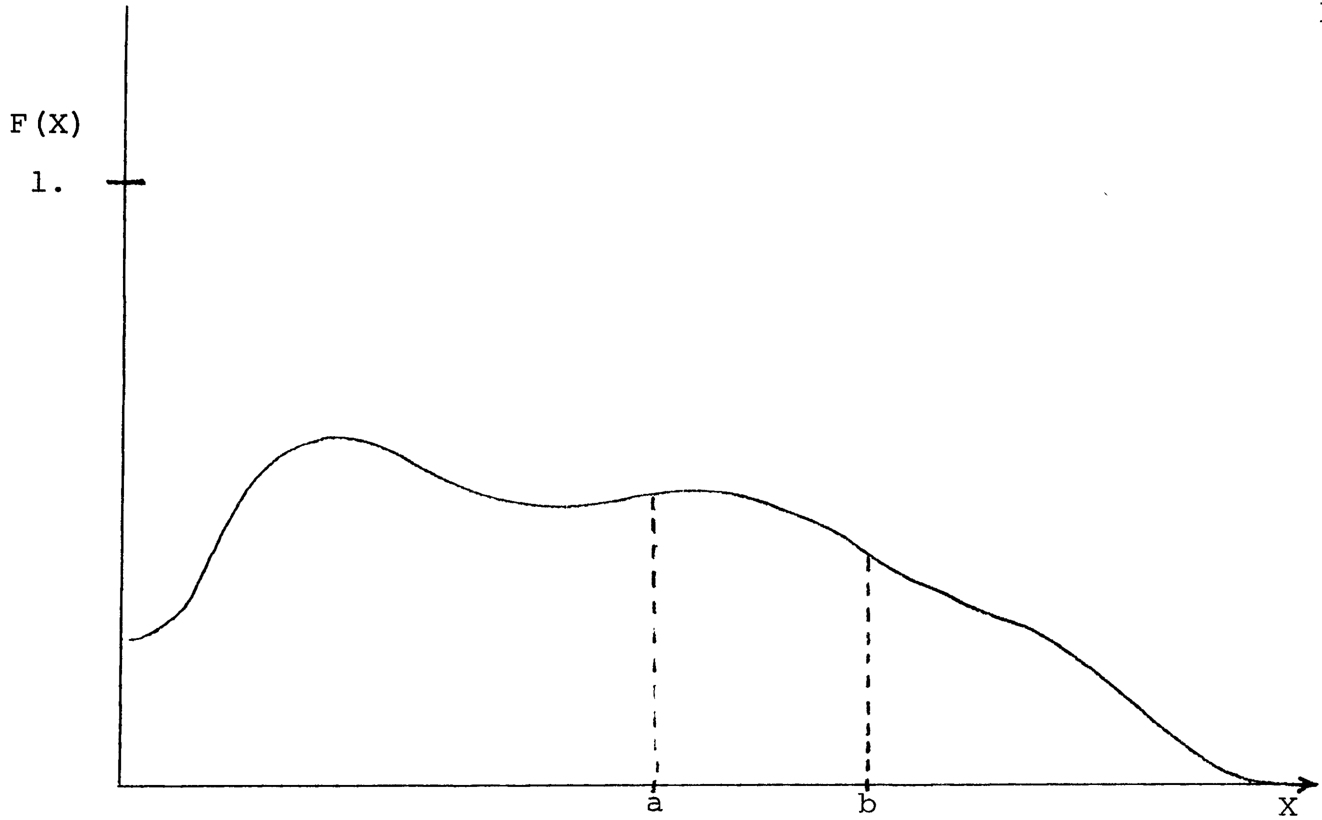


Figure 6.1. Example of a Continuous Probability Density Function

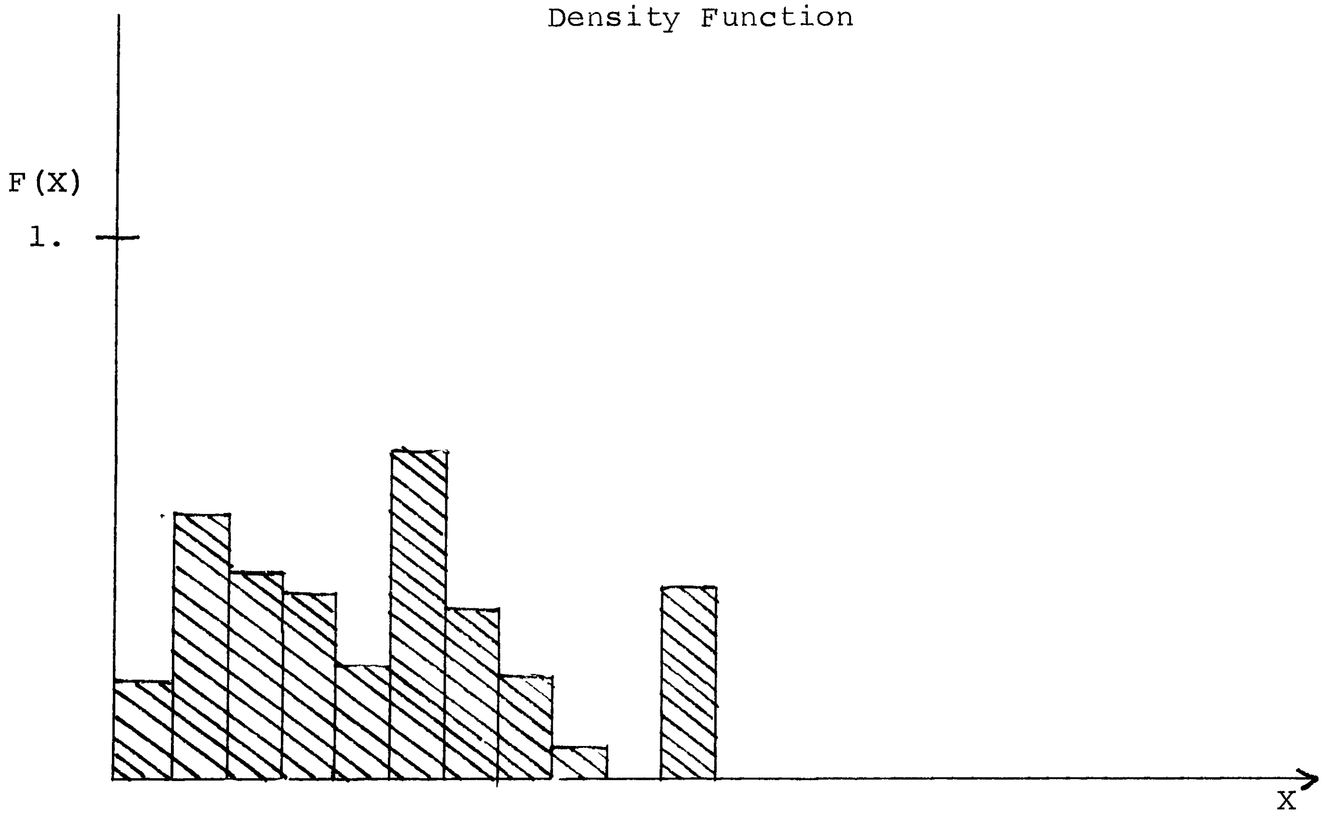


Figure 6.2. Example of a Discrete Probability Density Function

modelled by a complicated process which contains random effects. The time from take-off to landing approach is a function of meteorological conditions, aircraft type, air traffic control factors (e.g. altitude, separation, etc.), and other random factors. The time from landing approach to arrival at the gate can be modelled as air traffic control delays, ground speed and glide slope, and time from runway to gate. As can be seen from this example, even relatively simple processes can contain many stochastic events, of which some are permissible to be modelled as deterministic events, and some of which need to be modelled as random processes. The actual details of the modelling of the process can range over all degrees of complexity. It is the generation of these random events that makes a simulation model a Monte Carlo program. This method is the building block for a vast numberof models. Events of all types can be generated by Monte Carlo techniques. In doing any simulation, one must be certain that enough samples (or runs) are made that stable (or at least stochastically defined) results exist.

Queuing models have been developed for vaious sections of air traffic control.* Under certain circumstances, such as transient behavior or near capacity operations, the assumptions of most queuing models can be violated. An alternative approach is to generate events in real time through a Monte Carlo simulation. A model of this type can easily generate a time history of the system under many different scenarios and under the face of uncertainty. A similar type of model can be used for determination of the necessary capacity and frequency for a given route or system. Such a model potentially has many advantages over present fleet assignment

---

*Queuing models describe the waiting time that a user is subjected to before being served by some process (e.g. circling an airport before being allowed to land) in a precise mathematical closed form solution.

models that are based on math programming, but might be limited by execution time. These techniques could be used, with considerably more effort, for fleet planning or dispatching models. They have the capability of modelling the effects of aircraft breakdowns, weather delays, and other interruptions of service that math programming models cannot explicitly handle. It is not clear that the effort necessary to create these models would be worthwhile given all of the other conditions of uncertainty existing in an airline system. Surely models of this type will have their place in the future.

Simulation models can be very powerful tools for forecasting passenger demand. Large and highly complex simulation models can be created in a block by block process under a structured environment. As they cannot be divorced from econometric techniques, a modeller needs to have a firm grounding in econometrics. Although it would be difficult to prove so rigorously, simulation models theoretically have a greater potential for accuracy than purely econometric techniques. Even disaggregate techniques can suffer from problems such as multi-collinearity, heteroscedasticity, and identification. The limitations of simulation models are data requirements and their required interfaces with econometric methods.

Air transportation planning is rapidly moving from solely the realm of economics to a more integrated approach of management science and engineering. No longer can a modeller work in a totally econometric environment. It is only natural that the simulation techniques of science and engineering be incorporated into the air transportation modelling framework.

Reference 33 referred to this new class of models as transportation user behavioral simulation models (TUBSIM). These are disaggregate techniques that are based on the behavior of individuals rather than on statistical relationships that may or may not be able to explain market behavior over some limited time period.

The hypothesis upon which this particular behavioral simulation model is built is that a consumer makes his choice of alternative in a two-step process; 1) reject any alternatives not meeting <u>all</u> of the minimum utility (or maximum disutility) requirements; 2) of the remaining acceptable alternatives, choose the one with the greatest utility (or lowest disutility). In this particular formulation, step two is deterministic. It could easily be defined in a manner similar to logit for non-rejected alternatives:

$$P_i = \frac{e^{U_i}}{\sum_{\forall j} e^{U_j}} \qquad (6.2)$$

P_i = Probability of choosing alternative j

U_i = Utility of alternative i

Using the probability defined in equation 6.2, a given passenger could choose a mode in a stochastic (probabilistic) manner. In addition to the generation of additional random number required for a stochastic step two, more samples will be necessary in order to have a stable prediction. A small computer budget required using a deterministic approach for step two. Clearly a stochastic approach more closely resembles the true process, and is therefore superior.

Examples of this two-step selection process are present throughout all walks of life. When a person buys a tube of tooth paste, he might place a high value on cavity prevention, and will therefore choose toothpaste on the basis of cavity prevention provided that the best cavity fighter meets his minimum requirements for taste and whitening. A person flying from A to B might not give any substantial value to schedule delay because he can plan his day around it. However, he might require sufficient frequency so that he can be guaranteed of a morning or afternoon flight.

This two-step process can be used for trip generation, distribution, and modal split. For some purposes, it is necessary to define an alternative as "no trip." For this exercise, an intercity modal split model was developed. Considerably more work would have been necessary to include generation and distribution in a simultaneous structure along with modal split. Data and cost requirements dictated that only a modal split model be developed. The discussion following will be based solely on a point-to-point modal split model. Trip generation and distribution are just logical extensions of this model.

It is necessary to define all relevant alternatives for the market regions in question. Any level of detail and differentiation can be used, e.g. air can be considered an undifferentiated and homogenous service, or it can be broken down by class of service, carrier, non-stop versus multi-stop versus connections, time of day, or type of aircraft. The possibilities are limited only by the model's purpose and available funding and data.

For modal split models, the relevant trip characteristics consist of the factors that define level of service: total time, total cost, frequency, comfort, convenience, safety, etc. For direct generation models, other quantities such as alternative destinations, route time of day, and the "no trip" option need to be included.

The rejection criteria need to be defined. They may be simple requirements such as the maximum amount of money that a consumer would be willing to spend on a trip. They also can be joint conditions such as the minimum frequency required given some range of trip time. These can vary tremendously from consumer to consumer based on socio-economic factors, trip purpose, and other unexplainable reasons. The simple rejection criteria employed in the prototype model obviously do not conform to actual decision-making. However, since they are looser constraints than the actual ones, their inclusion can only help, not hurt the model. It is best to test all reasonable rejection criteria. If they are not significant, the mode choice will not be constrained by them.

The form of the utility (or disutility) function needs to be decided on. A linear or log-linear function is the easiest to implement and to estimate the parameters of. Non-linear utility functions can be used but require significantly more effort and time to formulate as well as being more difficult to estimate. Theoretically. discontinuous and non-linear utility functions can be developed that completely and accurately describe consumer behavior. Such utility functions could eliminate the need for the first step of the simulation process. As a practical matter, the present state of the art only allows for the estimation and calibration of simple forms for the utility function, thereby requiring the first step.

Much evidence indicates that there is a high degree of correlation among mode rejection criteria, characteristics of relevant alternatives and components of the utility function. In inter-city markets. bus and train users tend to be from the lower income classes and also tend to live in the central city which results in their having low access/egress costs. A consumer who values travel time highly. probably also tends to value waiting time and schedule delay time highly. The number of inter-relationships is large, but in most cases. only a small number of interrelationships is important. These correlations can be incorporated into a simulation model by segmentation of the market by trip purpose and socio-economic characteristics, by factor analysis, or by combinations of the two. Factor analysis is a technique for determining the underlying factors behind a process. For instance, one might include GNP and personal income in a model which could cause problems associated with multi-collinearity. Factor analysis might enable one to identify a single measure of prosperity as the underlying factor behind GNP and income. As disaggregate data is available, it is theoretically possible, although impractical, to directly generate behavioral units with all of the desired characteristics. Whatever scheme is used, it is necessary to generate passengers (behavioral units) that are identifiable such that trip characteristics, utility functions, and rejection criteria can be generated in some Monte Carlo method.

These techniques were applied to a model of modal split in the New York-Boston market. As the necessary disaggregate data was unavailable, it was necessary for the authors to make a small survey which will be described later. It is recognized that the data base that was generated from this survey has several statistical flaws. As the purpose of this exercise is to demonstrate the technique rather than to actually make forecasts, this problem can be overlooked.

The market was segmented into six categories based on trip purpose and socio-economic characteristics. The categories are: 1) single-day business trip, 2) multi-day business trip, 3) single-day personal trip (income > $10,000/year), 4) single-day personal trip (income < $10,000/year), 5) multi-day personal trip (income > $10,000/year) and 6) multi-day personal trip (income < $10,000/year). It would have been very desirable to segment the market into at least twice as many subdivisions, but lack of data disallowed this. It is expected that finer segments would have resulted in a model with better predictive powers and greater responsiveness to changes from the conditions of the calibration year. The large segments that were used in the prototype model result in losses of information due to aggregation. It is suggested that one not undertake any serious studies with market segments as gross as these. Finer segmentation could not have resulted in a poorer model (assuming sufficient data was available), only more work. If after estimation it was found that two or more segments had essentially the same characteristies, they could be combined into one segment if this relationship was expected to continue.

For this model, the alternative modes were air, bus, auto and rail. It would have been desirable to have finer differentiation of modes: several air fare classes, turbo and standard train, single occupant auto, car pools, regular bus, express bus, etc. An airline planner might wish to divide air into standard coach, first class, and each of the discount fares available, but to leave each of the other modes undifferentiated. As this is an abstract mode model, any new mode can be defined by its characteristics and added without restructuring or re-estimating the model. An abstract mode model is one in which the utility of a mode is calculated from the characteristics of the mode, not the mode itself. This precludes the use of mode specific variables. The theory behind this is that the consumer buys a bundle of attributes such as time, cost, frequency and safety without regard to what the mode looks like. He does not care whether he flies, travels on rubber wheels, metal wheels, or travels by an air cushion vehicle, so long as he travels on the mode with the highest utility. An outgrowth of this concept is that brand loyalty factors among various airlines is small. Although flaws exist with the abstract mode concept, it should be better than allowing large amounts of variation to be explained by the name of the mode rather than by its quantifiable characteristics. This model can handle the blue bus/red bus problem that is described in reference 34 which is similar to the situation where several essentially undifferentiated air carriers operate in the same market. In the red bus/blue bus problem, it is seen that for some models, if bus and auto each receive half of a market, if half of the buses were later painted blue and half were painted red, thereby resulting in three modes rather than two, blue bus and red bus would each get one-third of the market.

The level of service characteristics that were included in this model include door-to-door time, door-to-door cost, schedule delay, (which is defined as a function of frequency), and CC (a variable that describes consumers' subjective estimations of comfort, convenience, reliability, fear of flying, availability, mode or carrier loyalty, and other intangibles that was forced onto a one to five scale). Schedule delay time was considered separately from trip time because it was hypothesized, and later shown, that consumers value the two time differently. It would have been desirable to include many other characteristics such as breaking trip time and cost into their components, disaggregating CC, and including factors such as non-travel-related costs. For some trips, travel cost is a significant part of total cost; for other trips, it is a much smaller percentage of total cost, and is therefore of less concern. Although people's concepts of the meaning of CC differed, they were internally consistent.

The values of the level of service characteristics for each segment and each mode were characterized by a mean and a standard deviation that were computed from admittedly non-uniform and non-random samples. These values were generated by considering each behavioral unit to be characterized by a Gaussian distribution.*

---

*A Gaussian distribution has a single peak, is symmetric about this peak, and rapidly decreases in size as values depart from the peak.

In a real world, the actual distributions might be skewed, multimodal,* or thick-tailed. Segmenting the market in some way accounts for these problems, but does not completely solve the problem of generating characteristics with a Gaussian distribution. Limiting these and other values to a plus or minus three standard deviation range did not significantly affect the results.

Single rather than joint rejection criteria were used. These were generated from a Gaussian distribution with means and standard deviations computed from the survey. Conservative estimates of psychological bias were added to these numbers in order to correct the problem of using conjectures rather than actual actions in filling out the surveys. Adding in the bias appeared to improve the results.

The segmentation proved to be worthwhile. There were significantly different level-of-service values, disutilities associated with these level-of-service values, and significantly different rejection criteria among groups. There were no surprises in these values. For example, business and upper income personal travellers valued time more highly than did low income personal travellers; those making single day trips valued time and frequency of departure more than those making multi-day trips. In case and case again, it was observed that market segmentation revealed many of the intergroup differences. As previously mentioned, segmentation resulted in tighter distributions with better correlations among the model's parameters and input values.

As this model is disaggregated, it makes no sense to estimate the values of parameters of the utility functions with an aggregate technique. As the logit model (which was described in the previous chapter) has a linear utility function, is compatible with TUBSIM in other ways, and is also relatively inexpensive, it was used to esti-estimate the model's parameters. The output of the logit model

* Without a single, distinct peak.

consisted of the estimated parameter values, their standard errors, and their t-statistics as well as additional data that was not incorporated into the TUBSIM model. Socio-economic and mode specific variables were not used in the logit models. Instead, separate models were estimated for each segment. This assumes that each segment consists of behavioral units that are homogeneous except with respect to the available alternatives. This assumption is questionable for this particular implementation, but is probably acceptable for finer subdivisions of the market. If this objection is still not overcome even with very fine market subdivisions, it will be necessary to modify the simulation process to include some socio-economic variables.

Data problems were ever-present. It was necessary to estimate by educated guesses the percentage of the total market that each segment made up. Normally, the user would have another model available to predict this for forecast years. Enough data was collected so that the model would perform well. However, the quality and quantity of the data were poor relative to that required for most predictive purposes. For this reason, detailed reporting of the calibration and check-out runs will not be done. However, Figure 6.3 describes results from a previous TUBSIM model that did not segment the market. It was found that at least two hundred observations were necessary to obtain good results from each of the logit models. In several cases, not enough data could be collected. Attempts to collect additional data were not completely successful. Therefore it was necessary to eliminate some of the parameters for some segments since they could not be reliably estimated. For the most part, this problem did not affect the model development.

The following questions were asked on the surveys. They applied to each trip taken in the recent past and they applied to perceived values not true values.

Figure 6.3(a)

Description of TUBSIM Runs

Following is a description of the various runs attempted.

1) Base case to compare against results of the survey from Reference 39.
2) Base case with a change in the random number generator.
3) Base case doubling the number of iterations.
4) Base case quadrupling the number of iterations.
5) Base case quadrupling the number of iterations and changing the random number generator.
6) Base case increasing by a factor of eight the number of iterations.
7) Base case increasing by a factor of eight the number of iterations and changing the random number generator.
8) Eliminate the air discount fare.
9) Increase air fare by eleven percent.
10) Increase air fare by fifteen percent.
11) Decrease air fare by twelve percent.
12) Decrease air fare by fifteen percent.
13) Build a STOLport: psycar = min(0, psycar - .5), tbair - tbair - .1, the mean and standard deviation of access/egress time and cost decrease.
14) Congestion in the future: tbrail remains the same, tbauto and tbbus increase by .5 hour, tbair increases by .35 hour.
15) Rail fare increases by $2.50 which would eliminate part of the explicity and implicit government subsidy to rail.
16) Advertise air more heavily: psycar = min(0, psycar - .5).
17) Double air frequency.
18) Halve air frequency.
19) Increase road tolls and gas price: fareaut = fareaut + $2.00, assume that bus operators absorb the increase and keep fare levels constant.
20) Speed limit is reduced from 70 mph to 50 mph and gasoline prices remain constant: tbaut = tbaut + .5, tbbus = tbbus + .8, (assumes that buses observe limit closely, and automobiles do not).

Figure 6.3(b) Results of TUBSIM Runs

The following are the aggregated results of these runs.

MARKET SHARE IN PERCENT*

CASE	AIR	BUS	RAIL	AUTO
Survey Data	33.0	7.0	13.0	47.0
1	37.1	5.2	14.5	43.2
2	37.3	4.4	15.5	42.8
3	36.9	5.5	14.3	43.3
4	36.8	5.5	14.7	43.1
5	37.9	5.1	14.5	42.5
6	37.3	5.3	14.7	42.7
7	37.9	5.0	14.7	42.4
8	31.5	5.8	16.1	46.5
9	14.6	6.8	21.0	57.5
10	11.9	7.0	21.9	59.1
11	54.4	5.1	10.3	30.2
12	56.5	5.1	9.9	28.5
13	38.6	5.2	14.0	42.1
14	44.9	4.3	20.5	30.3
15	38.7	8.3	7.5	45.4
16	37.7	5.2	14.8	42.2
17	37.3	5.2	14.5	43.0
18	36.6	5.2	14.7	43.4
19	38.2	5.4	16.0	40.3
20	46.8	5.1	21.9	27.1

*Numbers might not add up to 100% due to roundoff.

1) For the chosen mode and all alternative modes, list your perceived values for door-to-door time, door-to-door cost, frequency of departure, and CC (note: CC, which is the comfort/convenience variable, was explained in detail to each respondent).

2) Under normal conditions, what are your minimum requirements for time, cost, frequency and CC?

3) Identify what market segment you're in.

Most people were able to give multiple responses as they have fallen into several categories at various times. Respondents were asked to answer not only for themselves, but for anyone else whom they could confidently answer for e.g. friends, relatives. In order to obtain enough data for the logit models to converge to reliable values, respondents were asked to list their responses for trips in their recent memory rather than for just their last trip.

Questions were asked about individual's rejection criteria and about the perceived characteristics of the chosen and all alternative modes. For all segments except the lower income groups, complete availability of all modes was assumed. For the lower income groups, automobile was the only mode that was potentially unavailable. There were difficulties in defining auto availability. Many people could have ridden in car pools such as those arranged by university ride boards and local radio stations. This resulted in difficulties in defining schedule delay, comfort/convenience, and availability. There was tremendous variance in this part of the data. Theoretically such car pools were always available with uncertain schedule delay and costs and with poor comfort and convenience. As a practical matter, these car pools could often be considered unavailable.

These statistical problems undoubtably bias the estimates. It was decided to accept any parameter that had the correct sign and a t-statistic that was significant at the ninety-five percent level.

Despite the data problems, the parameter estimates appeared to be "in the ballpark." For business trips and single-day personal trips for the upper income class, a large majority of the sample took air. As air was the most expensive mode, this made cost appear to be a utility rather than a disutility. Even collecting additional non-air data did not result in significant t-statistics for cost for single day business travellers.

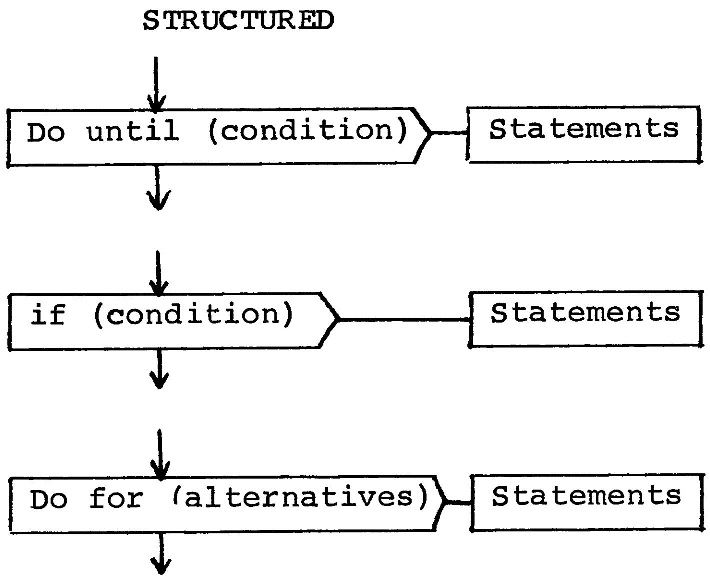
Almost all disaggregate data bases will contain some inconsistent observations. If the survey was obviously misunderstood or answered ridiculously or incompletely, those observations should be discarded. However, there will still be inconsistent observations that result from an incomplete or misspecified model or a minor misunderstanding of the survey. An example of an inconsistent observation would be an individual who chooses a mode that is poorer than one or more of the other modes in all categories. For a sufficiently large sample, the effect of a few inconsistent observations should be small. For the small number of samples used in this exercise, inconsistent observations biased the coefficient estimates and greatly increased uncertainty.

If one were creating a TUBSIM model whose purpose was predictive rather than illustrative, it would be necessary to collect a suitable data base. If it is assumed that two hundred observations per segment are necessary, that each household survey results in five observations (several different people and several different trip categories), and that twenty market segments are defined, at a cost of twenty to fifty dollars per household per interview, this would result in a cost of sixteen thousand to forty thousand dollars. This cost is within the reach of many airlines and government agencies. In addition, this data can be used for several different studies, and once a model is calibrated, it can easily be transferred to another market region. Once disaggregate forecasting techniques become more popular, it is likely that many disaggregate data bases will become public or semi-public domain.

In order to collect aggregate data bases, one must sample individuals. Future transportation surveys or government census's should retain the disaggregate data in its original form as well as aggregating it. It would be useful if even a small fraction of the original data were preserved in its disaggregate form. For such disaggregate data bases, the collection costs would not exceed those of aggregate data bases. The only additional costs would be storage and clerical charges which would be small relative to the collection costs. It is wasteful not to retain the data in its disaggregate form as long as one originally has it available.

The parameters of the utility function are estimated from observations of individuals, and therefore do not contain psychological bias and do not necessarily require a uniform data sample. The level of service characteristics and the rejection criteria are generated from individuals' conjectures, and therefore contain psychologoical bias and require a uniform data base. Conservative estimates of bias were added in to partially eliminate this problem. It was assumed that people underestimate the worst condition that they will accept. It was further necessary to constrain the modal characteristics to be non-negative and within normally encountered ranges.

The simulation program was structured into subroutines that can be altered with only a minimum of effect on the rest of the program. This structure makes it easy to make changes in specific modules that reflect differences in degrees of accuracty, detail, and level of aggregation. The flow chart (in structured notation) of the particular version that was coded is presented in Figures 6.5, 6.6, 6.7 and 6.8. A description of some of the differences between structured and conventional flowcharts is presented in Figure 6.4. A program



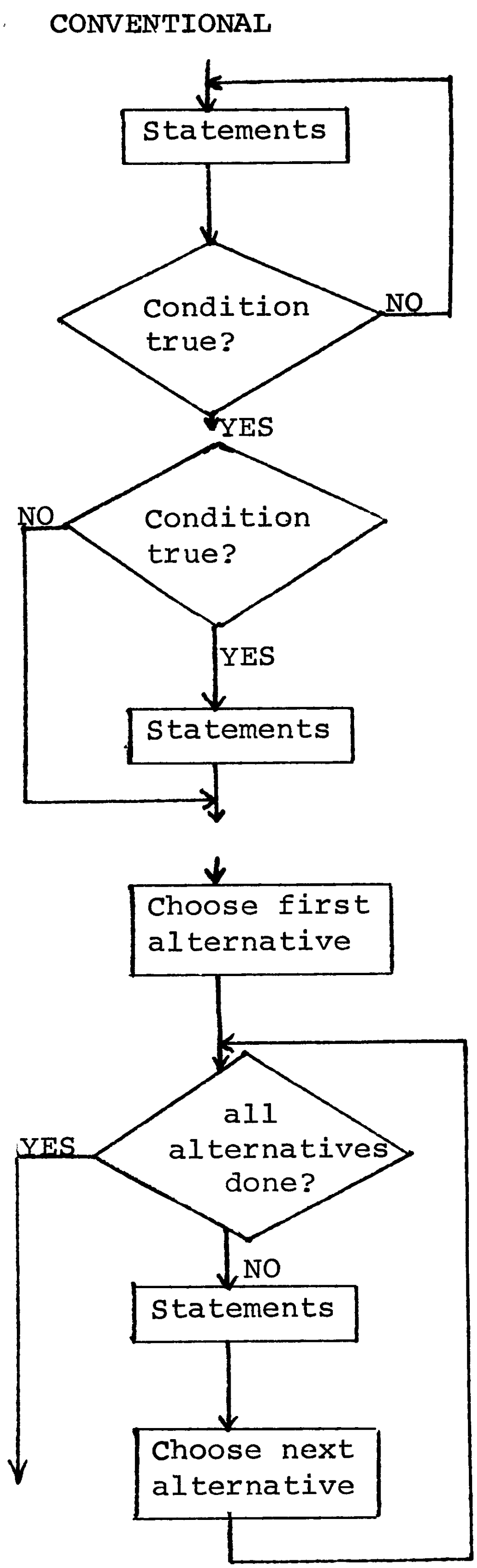


Figure 6.4.

Structured versus Conventional Flowcharts

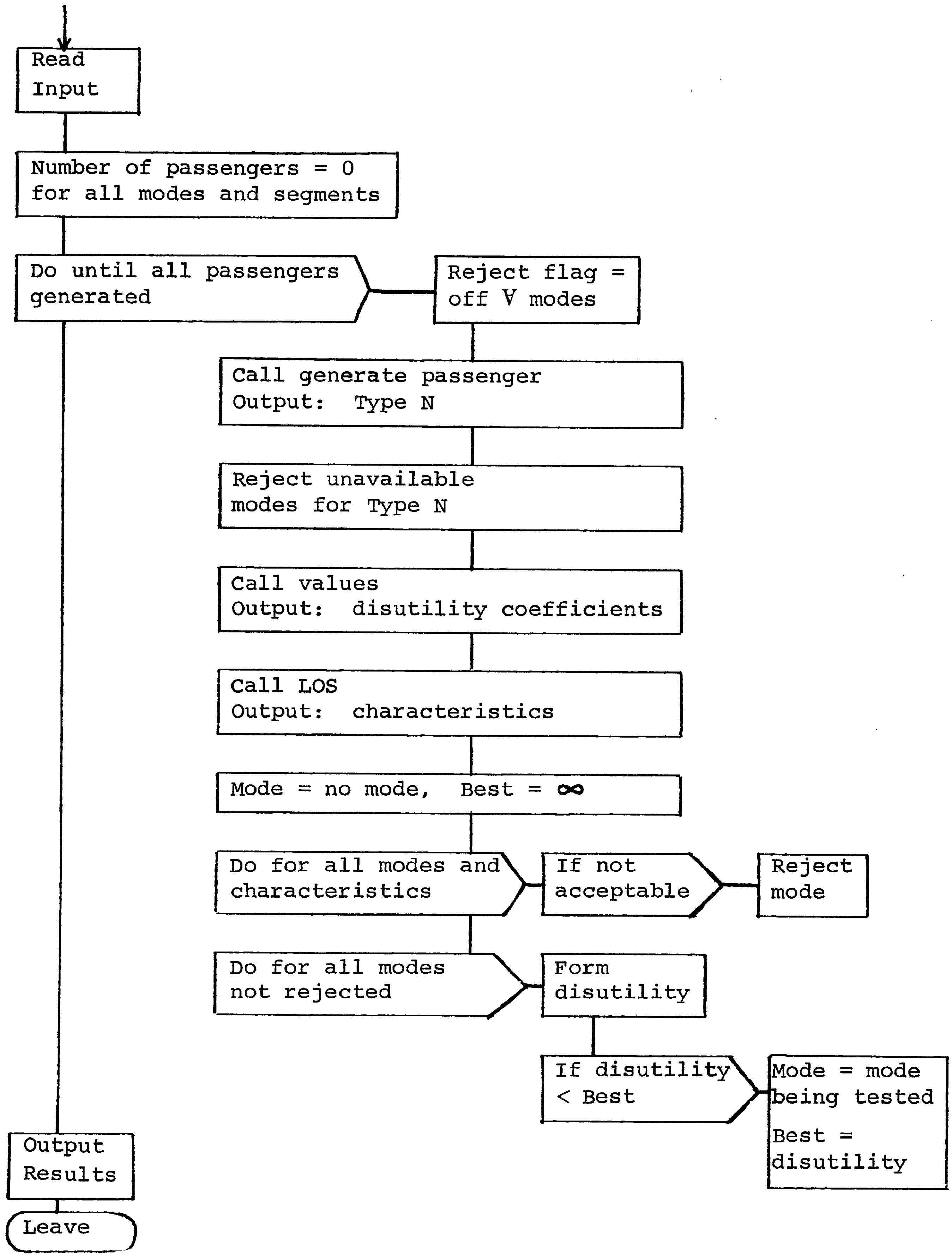


Figure 6.5. TUBSIM Outer Program Level

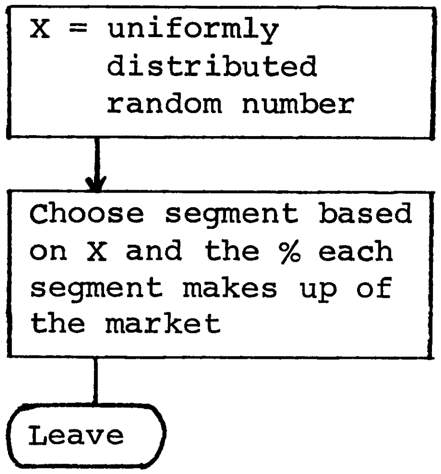


Figure 6.6. Generate Passengers

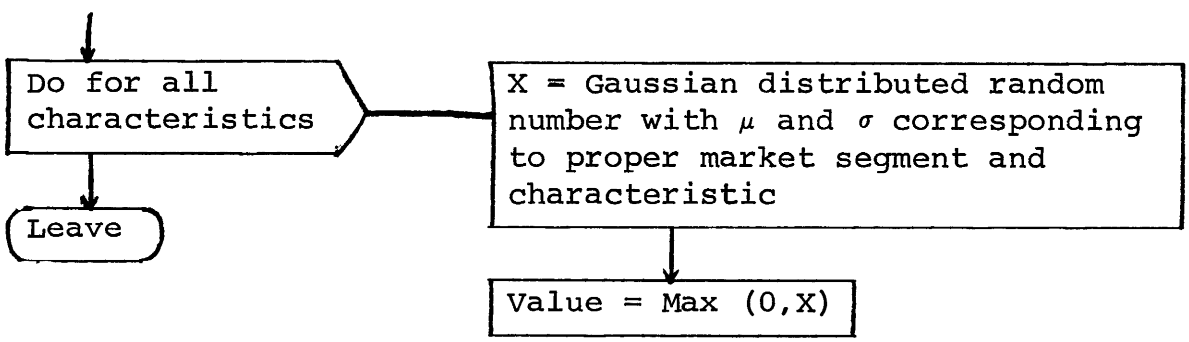


Figure 6.7. Values

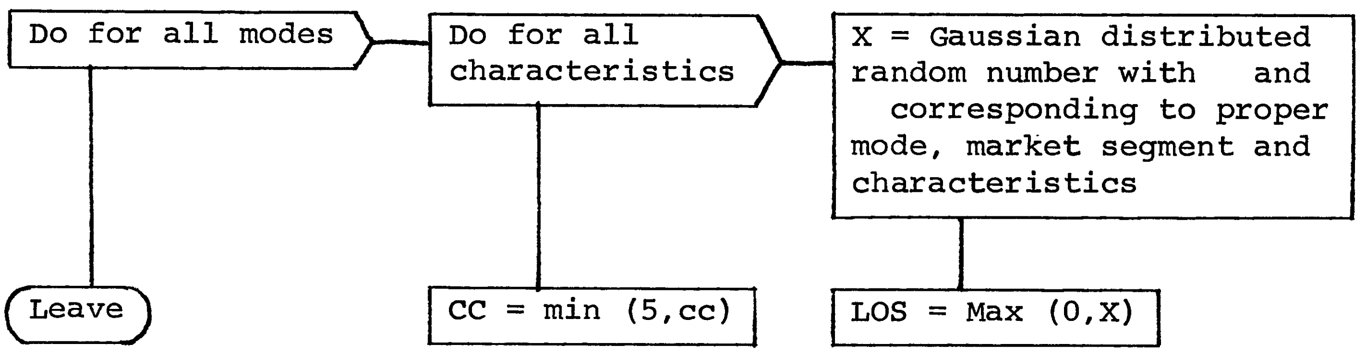


Figure 6.8. Level of Service

listing appears in Appendix D. This version of TUBSIM is very simple and straight-forward in order to make it easy for others to understand and also to be able to meet computer budget constraints. This basic framework can be built upon to any degree of sophistication that is desired.

Following is a description of the outer level of program flow.

1) read inputs
2) initialize the number of passengers choosing each mode to zero;
3) do for as many passengers as desired;
4) initialize mode rejection switches to off;
5) describe passenger, i.e. generate market segment
6) reject all unavailable modes;
7) generate rejection criteria;
8) generate values of coefficients for the utility function;
9) generate trip characteristics;
10) initialize: best mode = no mode, best disutility = infinity;
11) reject any mode that does not meet all rejection criteria;
12) compute the disutility for all relevant (available and not rejected) modes
13) choose the mode with the lowest disutility (alternatively choose mode in a stochastic manner based on disutility values);
14) accumulate and output data.

The outer structure is relatively simple and complete. This structure with minor modifications is probably suitable for most problems that would be encountered. A high degree of personalization and flexibility is allowed for in the design of the inner modules. The version that was coded used matrix notation which resulted in greater efficiency at the price of some loss of visibility. From working with this model and a similar one from Reference[33]. it was generally found that as the number of random factors increases, the number of runs required to obtain stable results increases. For this model, stability is defined as the point where increasing the number of passengers generated or changing the sequence of random numbers does not result in significant changes in the modal split. For this particular problem, at least three thousand iterations are required for stability to occur. Increasing the number of iterations was found to have decreasing returns to scale with respect to the model's stability.

The model described in Reference 33 used discrete rather than continuous values for reject criteria, coefficients of the utility function, and level of service characteristics. If used properly, discrete generation can result in a more powerful tool than continuous generation under certain circumstances. When working with a discrete formulation, one needs to ensure that quantization effects do not degrade the results. Highly non-linear situations such as almost no sensitivity to change occurring over some dead band, and then almost instantaneous change elsewhere can occur in discrete formulations.

For this model, approximately five percent of the sample rejected all modes. This figure was approximately half of the number of rejections that were experienced in a non-segmented model. Some rejections occur because the generation process does not limit the values to some reasonable range of values such as plus or minus a two or three standard deviation range. Finer market segmentation would have eliminated many of those consumers who rejected all modes by resulting in better correlation among the model's parameters and coefficients. In actuality, a person who places a high value on time is usually willing to pay for it. A person who cannot afford to pay very much realizes that he has to take a slower mode that is probably less convenient and less comfortable. The version of this simulation model did not take account of this except for the effects of segmentation. Increasing the correlations will significantly increase execution time, but will also eliminate much of the problem of a consumer not being able to choose any mode.

The final forecast of a model should be the predicted shares rather than just the predicted individual probabilities. One can estimate the modal shares directly from the data used to calibrate the logit models.

$$E(i) = \sum_{i=1}^{T} P(i:A_t) \tag{6.2}$$

E(i) = expected number of consumers choosing alternative i out of a market containing T consumers

This procedure is not always strictly valid because the data base might not be appropriate due to its usually small size or lack of sufficient care in collecting it. It was seen that an over-abundance of air travellers for certain segments required the collection of observations of non-air travellers in order to estimate the disutility of cost. As no mode-specific or socio-economic variables were included, the coefficient values probably were not significantly biased. However, the estimated modal shares will be. Even without this problem, it is unlikely that one will have available the statistics of alternative trip characteristics and socio-economic characteristics for a different market region or for a forecast year.

Aggregation can occur at this level. In some instances, aggregation can seriously bias the modal share estimates. If possible, some disaggregation, even if it is just business travellers versus personal travellers, should be used in the hope of improving modal share estimates. Even with aggregation at this stage, one can expect potentially better results than from a model estimated entirely with aggregate data.

$$E(i)_{aggregate} = T \cdot P(i:A_{av}) \tag{6.3}$$

A_{av} = alternatives available to the average behavioral unit

The simulation technique that is presented represents a feasible method of taking disaggregate probabilities and estimating market shares. The model was stable and behaved as expected. A large amount of sensitivity analysis and predictions for various scenarios such as gasoline price increases with lower speed limits, and the building of STOLports was done. Numerical results are meaningless due to the questionable data base used for calibration and due to the lack of generation and distribution models to "feed" the modal split model.

Aggregate elasticities for small perturbations of air travel time and air cost were computed by varying the values of the parameters and observing the results. These numbers were in the same range as elasticities computed from aggregate models. Market segmentation and estimation of coefficients using logit appeared to improve the model relative to the more elementary model described in Reference 33, but once again, data limitations made it difficult to evaluate the results. Perturbing the inputs demonstrated that no particular quantities were especially critical except for perhaps the percentage of the market that each segment makes up. This was an illustrative rather than a numerical exercise; and as such, served its purpose.

The TUBSIM model with disaggregate estimation of coefficients allows for inclusion of many minor effects that would cause statistical problems in aggregate models. A linear utility function was used only because of computational simplicity and compatibility with the logit model. There are no other reasons not to use a non-linear utility function if a priori reasons dictate it.

This model does not require demand elasticities as inputs, but can generate them for use in policy analysis. There appear to be no fundamental limitations on uses of this model. Most of the statistical problems that are encountered in econometric models are eliminated in a TUBSIM model. It can be used for evaluation of presently uncertain processes such as what percentage of refused reservation requests are recaptured on another flight, and what percentage divert to other carriers or other modes; modelling of inter-carrier competition; the introduction of a new type of fare or a new mode; and the effect of switching a flight time, ad infinitum. Little or no recalibration is required for many problems. As this is an abstract mode model, it can model many situations that other techniques cannot handle.

This technique is straight-forward, simple, and is based on behavioral assumptions. Its structured form allows for it to easily be moved to different regions and to be modified in a modular simple manner. These qualities make it very adaptive. Computer time requirements can become large, for very sophisticated implementations, but still are not excessive. The modules can be developed independently by experts who might be familiar with only one or two modules. The program is designed so that the outer structure can be created by management without the requirement for detailed analytical knowledge, and the inner structure can be designed by technical personnel.

Almost any level of aggregation can be built in. Supply effects, dynamics and other feedback relationships can be added with little effort. The data requirements are low relative to those of aggregate econometric models. However, data is not yet available for most air transportation problems of interest.

The human behavioral relationships that are modelled in a TUBSIM model are more likely to remain constant or at least predictable over time than aggregate statistical relationships; i.e. the aggregate structure can change without the individual behavioral structure changing. This can result in a longer predictive horizon. This technique can be used for generation, distribution, and modal split in either a recursive or a simultaneous formulation. It has many advantages over aggregate econometric techniques with very few disadvantages.

The class of model just presented is only one of the many simulation structures that is possible. Although infrequently used at the present time, simulation techniques are useful and will probably grow in importance for transportation demand forecasting.

Chapter VII

Model Evaluation

The importance of selecting the appropriate technique for forecasting a particular aviation activity can not be underestimated. An inappropriate technique no matter how sophisticated can result in an inferior forecast. There are a number of criteria which the analyst can use to evaluate various forecasting methods. However, there is still a certain amount of judgment required in selecting a method which is best suited for a particular aviation activity. One thing is clear, it is not always true that the overriding criterion should be accuracy. It is generally a trade-off between a number of criteria. The analyst has available to him a set of general guidelines to select a technique to suit his specific situation. In this chapter, we discuss some of the more important issues which are relevant in evaluating forecasting models.

A model should attempt to capture the fundamental relationships of the process being studied. This does not necessarily mean that the true explanation of the process needs to be quantified. A chartist claims to be able to capture the fundamental relationships by observing past patterns; a marketing manager captures the relationship by observing consumer sentiment either from surveys or by talking with salesmen; and a sophisticated analyst might actually attempt to write down a set of mathematical equations that he believes describe the process. Models that have a greater relationship with the real world quite often are more easily understood than others due to the closeness of them with the user's intuitive subjective models. This quality makes them easier to work with.

Any model that is used should be policy sensitive. A travel demand model of North Atlantic travel calibrated over the past ten years could perform excellently without aircraft speed as a factor. However, if one were attempting to estimate the effects of the introduction of a supersonic aircraft, this model would obviously be insufficient. Factors that seem unimportant now might become dominant in the future. The model's design should be consistent with its intended usage and fit the task at hand. Many models used for forecasting were designed to explain the structure of the problem. There is no reason to expect that complex, explanatory models will necessarily be suitable for forecasting, and vice-versa. No one model is best or even suitable for all purposes. A model should be sufficiently good that one is able to use it to make realistic and firm recommendations (both structural analysis and prediction).

Data requirements are of utmost importance. The best model in the world is of no value if the data necessary to calibrate and run it is not available. In such cases, it is far better to construct a model that is less accurate, but for which sufficient data is available. In evaluating data requirements, one might wish to consider availability, quantity, consistency, accuracy, format, and the cost in time, manpower, and dollars necessary to collect and maintain the data base. It is quite often possible to calibrate a model with historical data, but to be unable to use it for forecasting due to lack of projections for the independent variables. One might be able to obtain an excellent historical fit of air travel versus telephone calls or some measure of consumer confidence. It is doubtful that this model would be useful for forecasting as projections of telephone calls or consumer confidence would probably be unavailable or unreliable. This type of problem is particularly apparent in econometric models whose independent variables include the lagged value of the dependent variable. In such a case, one might wish to eliminate the unavailable variable or to take care of the auto-regressive term by setting up a multi-equation simulation model. An example of this is discussed in the econometric section.

Calibration is necessary before a model can be used. The ease of calibration depends on data requirements and model structure. Convergence and uniqueness of solution are also important issues. Calibration is usually achieved through slight manipulations of the data and the model structure, and perhaps through an iterative process. The data and structure used for calibration need to be valid for the entire forecast and calibration period. In general, for well-behaved data with no large perturbations off the trend and seasonal patterns, better calibration is achieved with larger data bases. In econometric models calibrated with consistent estimators, the mean square error of the estimator approaches zero as the number of data points increases to infinity. In other words, both the bias and the variance of the estimator become very small for large data bases. However, extending the size of the data base will not increase confidence in the estimates of the model's parameters if the estimators are not consistent or if the structure and/or data change during that period. As an example, consider the case of transatlantic travel. In recent years, the structure of the process has changed as charter activity and all-inclusive tours have increased. The data base has also changed as a different group of people are now travelling (more people from lower income classes). The elasticities estimated are now computed from aggregate statistics of different groups. There is no reason to expect that the demand elasticities of a 1962 group will be the same as the elasticities of a 1974 group.

The level of accuracy required in a model is highly dependent upon model purpose and the accuracy and costs of alternative models.

Every model used for decision making should be responsive to possible decisions of the user or of the human environment. However, this must be consistent with the real-world conditions. A model might show a linear relationship between frequency and volume; however, it is highly doubtful that volume will continually increase without limit as frequency increases. Sensitivity analysis should be used to point out the critical areas of the model and to impart some information on the amount of confidence that should be placed in a given model. Although highly accurate models are desired, any increases in accuracy must not be offset by an even larger increase in costs or other factors. Better and more accurate results are of no value if they do not contain additional information that can potentially change recommendations or actions.

Designing a new model for each new task is time and money consuming and represents duplication of effort. It is far better if possible to modify existing models to suit the particular problem being studied. Unfortunately this cannot always be reasonably accomplished. Unless it can be guaranteed that the model will be used only for the task at hand, it is desirable to have a flexible model. By this, it is meant that the model can be easily applied in different locations, by different airlines or agencies, and at different points in time. A model that is designed to predict New York to Boston air travel for a specific airline in 1974 should be able to be modified to predict traffic for a different airline on a similar route in 1975. An overly specific model without a good theoretical foundation should be examined closely for misspecification. The scatter plot in figure 7.1 might be reasonably forecast by predicting total passengers as a function of disposable income. The outliers can be handled by the inclusion of dummy variables; thereby

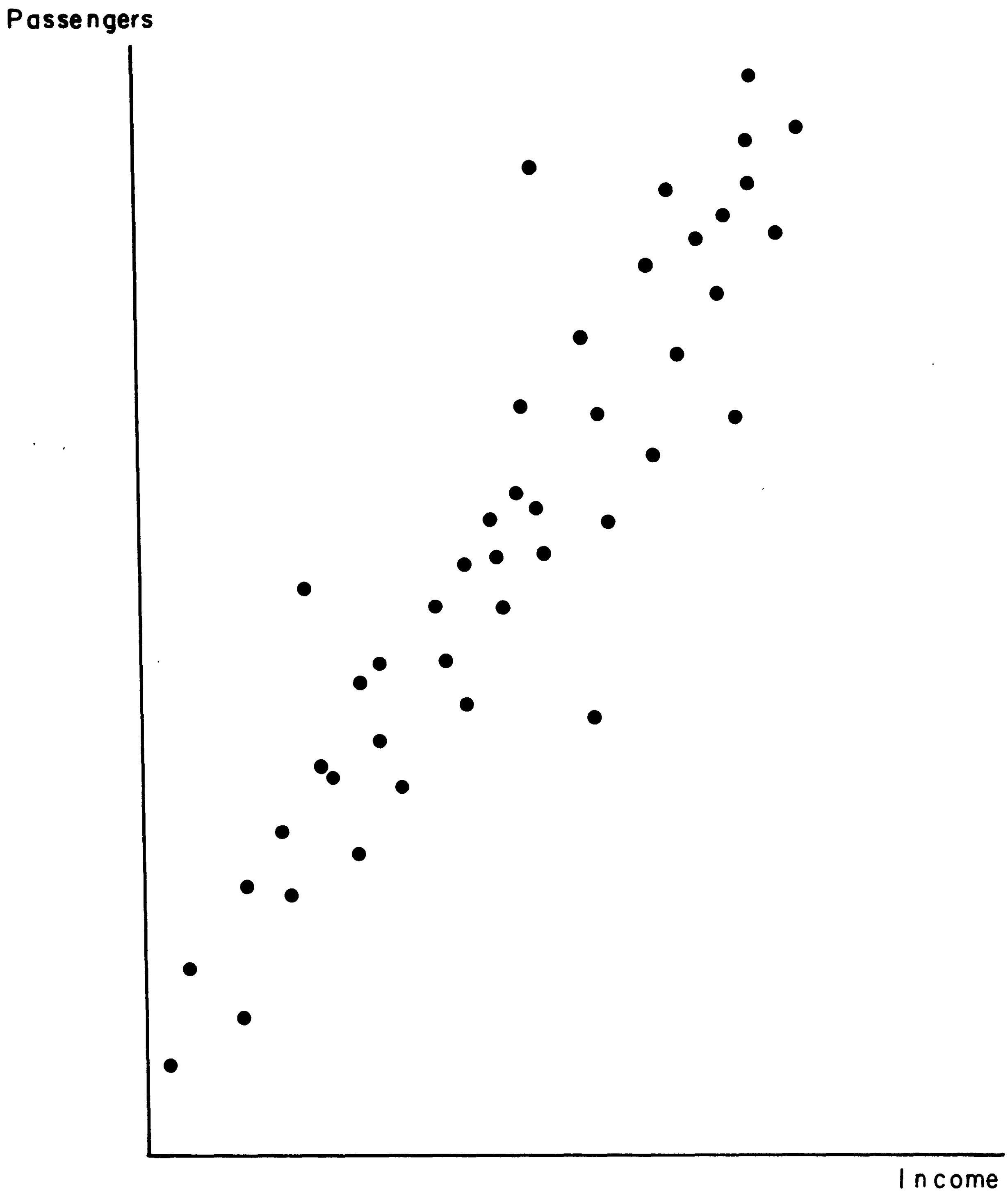


Figure 7.1 Scatter Plot

insuring a better fit. Although this model might be well suited for the particular data, it probably could not successfully be applied to other areas. Rather than achieving a good fit with meaningless dummy variables, it is better to accept the poorer fit or to try to explain the data by inclusion of other variables such as fare, travel time, or changes in other modes.

The cost of operation is of great concern to many organizations. This cost appears in maintenance of the data bank and model, actual running of the model, and evaluation. The costs stem from manpower requirements, computer costs, and other money requirements. A model should always be evaluated in terms of costs versus benefits.

Some models are capable of standing alone. Quite often, there is the need for several models to interact with each other or with other data bases or projects. This sometimes occurs in a formal structure such as there might be when land use, trip generation, trip distribution, and modal split models form a system. Interaction can occur in a less structured manner when a consultant builds a model for a client who already has a library of models. Problems can occur on the highest level or in such a seeming triviality as two models requiring the same data, but in different formats.

The issues discussed so far are very general and apply to all types of models in all categories described in this report. However, within the categories of intuitive subjective models, control theory (systems dynamics) models, statistical (econometric) models, and simulation models, finer subdivisions exist. Models can be disaggregate or aggregate, simultaneous or recursive, stochastic or deterministic, dynamic or steady-state (equilibrium), long, medium, or short term, and structural or predictive in nature. Distinction can also be made between direct and indirect models. However, this category is just a special case of level of aggregation. The distinctions among categories are not always clear, and various combinations can exist. Thus, model evaluation must now be undertaken at the finer level. Following are some of the examples of model evaluation at this level.

The end result of a forecast is usually a prediction of some aggregate measure such as total trips. An aggregate forecast can be made directly from an aggregate model or can be made by summing the actions of behavioral groups in a disaggregate model. Aggregation can occur on any level from an individual consumer to the total market. Oftentimes, the behavioral unit is assumed to be larger than the individual such as the family or all business travellers making single-day trips originating in a particular zip code area.

"If detailed observations are available, aggregation for forecasting could be performed either before or after the model estimation. Aggregation during the model construction phase of the analysis will cloud the underlying behavioral relationships and will result in a loss of information. It is always desirable to estimate a model at the disaggregate level."[36]

For some simple cases, aggregate models will yield estimates that are averages over the values in a disaggregate model. For more complicated and non-linear cases, this is not strictly true. The aggregate model would not account for non-linearities in behavior and would not account for skewness or width of a distribution (e.g. access/egress cost). Disaggregate models make use of variations within aggregation groups and among aggregation groups. Aggregate models only make use of variations among groups. This greater variation in disaggregate models often causes multicollinearity problems tc disappear. For practical purposes, disaggregate models always have more information available to them than do similar aggregate models using the same data base unless the data is fully aggregated initially. Disaggregate models tend to be more flexible and can serve different purposes more easily than an aggregate model because a disaggregate model can be aggregated on any level desired.

Aggregate models require data sampled from uniform cross-sections of a market region. Disaggregate models need only sample a large enough group such that many different behavioral units are included (300 interviews are often sufficient). Therefore, a much smaller data base is required for disaggregate models than for aggregate models. Furthermore, if it is assumed that the behavior of people in city A is similar to the behavior of people in city B, a model calibrated in A can also be used in B. The savings in data collection time and cost can be very significant. Also cited as advantages of disaggregate models are: 1) they tend to be more policy sensitive than aggregate models, and 2) they are suitable for sub-regional and project planning as well as regional planning.

Recursive models sequentially estimate various elements of a consumer's set of travel choices (frequency, destination, time of day, mode). Simultaneous models, as the name implies, estimate all elements of a specific trip or trip category at one time. In a recursive framework, one might decide that he needs to make a trip to city X, then he decides to make this trip twice this quarter, he then chooses air, and then decides what time of day to travel. His route choice depends on whether he takes a non-stop or multi-stop flight. This decision is made when he makes his time of day decision. A simultaneous model would consider all factors simultaneously. Clearly, the simultaneous structure best approximates a real-world decision. If a recursive structure is proper, a simultaneous model would just degenerate into a recursive structure. Recursive models assume a given priority in choice. Does a vacation traveller decide what mode to take and then decides where to go given that mode, or does he decide where he wants to go and then chooses the best mode? The answer is not obvious; and the assumed hierarchy of choice affects the forecast results. Unless computational, cost, or data requirements dictate the use of a recursive model structure, a simultaneous structure should be preferred.

The distinction between probabilistic or stochastic and deterministic models is clear from the names. Stochastic models assume some probability of making a given choice rather than complete certainty. Logit, probit, and tobit are examples of stochastic models. The expected value of a stochastic forecast is often just the deterministic forecast from another type of model. At the present time, deterministic models make up the bulk of the research in air transportation forecasting.

Consumer choice is usually defined as some function of socioeconomic, level of service, and other relevant variables. There are two primary examples where it is observed that deterministic models have difficulty explaining consumer choice. Identical consumers under seemingly identical situations have been observed making different choices. The same consumer in the same situation has been observed making different choices. This probably occurs because the models do not include all of the elements that define a possible situation. It would be impractical to do so. One would have to include the consumer's mood, what time he woke up, and anything else that could possibly affect his travel decision. In most cases that the authors are familiar with, the use of deterministic models did not seem to noticeably bias studies so as to make them unacceptable.

In many situations, the dynamic response of a system is significantly different than the steady-state response. For reasons of simplicity, most models deal with what is usually called an equilibrium solution (either short run or long run). From a thermodynamic point of view, the term steady-state is more appropriate than equilibrium. It is difficult to model the dynamics of a situation except with subjective models and control theory models. A partial modelling of dynamic effects can be done in simulation and econometric models by using variables that are functions of time such as lagged variables, compound growth, and exponential decay. There is little problem in doing this for simulation models. In econometric models, these techniques usually at least require more computer time for most regression packages, and in cases of serial correlation, they can result in biased estimates of the coefficients. For many purposes, the dynamic effects are long term and can be ignored in short term analyses. In predicting the expected demand at a proposed new airport, many short term effects can be modelled. There is a longer term effect in that a new airport will probably attract new businesses and hence new residents to the market region which would in turn increase airport traffic. For some analyses, it is appropriate to ignore this demand shift. If a hierarchy of choices and model types is assumed (e.g. generation, distribution, modal split), various levels of dynamic response can be modelled by "freezing" different levels of choices and considering dynamic response only at the remaining levels.

In air travel, most perturbations are small e.g. six percent fare increase, two percent growth in population, five percent increase in daily frequency, ten minute increase in access time, etc. Occasionally, there are large perturbations where dynamic response is probably important e.g. introduction of jets, introduction of the SST, non-stop service introduced for the first time, etc. One needs to be careful to include all important dynamic effects.

Models can be primarily predictive or structural in nature.

A structural model attempts to explain the actions of a system by showing cause and effect relationships. A predictive model is solely concerned with forecasting a dependent quantity. There is no reason for concern in a structural model if heteroscedasticity, serial correlation, poor fit, or other statistical problems exist so long as unbiased estimates of parameters are made. For the purpose of fare regulation, one might not be concerned with prediction of travel volume, but would want a reliable estimate of demand elasticity with respect to fare level. On the other hand, if one is only concerned with forecasting next year's traffic, it does not really matter what the value of fare elasticity is as long as the model has good predictive abilities. "The record to date does not reveal any significant advantage of the much more complex structured model over the 'seat-of-the-pants' judgemental-type model used by many business forecasters and business firms. But the fact is that judgemental-type models offer no hope for improvement, while structured econometric models offer the only real prospect for being able to do better in the future."[37] This quote from Juster refers to business forecasting. Already, the field of transportation forecasting is to the point that structural and behavioral models often yield better results than naive (although often complicated) predictive models. Presumably, these advances have been made possible by having the new point of view brought in by engineers added to the work already done by econometricians.

It is difficult to estimate structural coefficients for air transportation models because so many changes happen at once. A large problem exists due to the many fare levels and packaged tours that exist. A price variable is not uniquely defined. Coach fare or yield is not completely satisfactory as few people actually travel at exactly those levels, and significant aggregation problems exist. The minimum fare level should give some indication of total market size, but is unacceptable for the aforementioned reasons and because not everyone is eligible for the minimum fare. If the data were available, it would be possible to create a partially disaggregated econometric model which would have common information for many of the variables, but would take account of the differences in passengers travelling at different fare levels. If a detailed disaggregate data base were available, many of these objections could be overcome by using a logit or probit model. The cost of collecting and maintaining the required data base would by exorbinant with the present system and technology. For the time being, this improved method of forecasting is infeasible due to data requirements.

Also complicating analysis is the diversion that can occur in a complicated transportation system. Charging an individual fare will affect not only the volume of travel at the particular fare level, but will affect the entire market as well. If a lower fare class is instituted, not only will new passengers be attracted to the system at that fare level, but also the entire distribution of passengers will be affected as individuals downgrade their service and thereby affect the yield. It is important to consider the entire system. If fares to one vacation market are increased, some travel will divert to other vacation markets. As previously mentioned, it is also necessary to consider the dynamic nature of the situation. The impacts of fare changes, new competition, and other changes in the system do not happen immediately or all at one time. Further compounding the situation is the impossibility of running a controlled experiment. Many changes, some controllable and some not, are happening throughout the time period being studied.

It is often easier to predict specific (e.g. single city pair, individual carrier, etc.) data rather than aggregated system wide data. As the expected prediction becomes smaller, noise in the data and model misspecification becomes more noticeable. There is some critical level below which quantization and noise hurt predictions, and above which accuracy decreases. An analogy can be made to forecasting the GNP. It is more accurate to forecast individual components of the GNP and then add them together than to try to directly forecast the GNP. However, it is unreasonable to attempt to forecast every minute component of the GNP. This "critical size" is usually difficult to determine and fortunately is not often important as intuition usually correctly guides the forecaster.

There is always the lack of information to include everything of consequence in a model. At finer and finer levels of detail, the question of whether anything is forecastable becomes both psychological and philosophical. One is never certain of what is best to measure or how to measure it. For instance, is the relevant supply variable for a city-pair seats or frequency or both? Due to multi-collinearity, both should not be included. Using only one of these variables results in loss of information; the specification of a variable that includes both of these effects is not a simple thing to construct. The best that one can hope to do is to isolate the most important statistics and interactions.

In air travel passenger forecasting, one is basically concerned about three different types of models: air trip generation and/or attraction, air trip distribution, and modal split. Trip generation and attraction models can answer questions such as how many passengers will airport X handle five years from now. Trip distribution models, in particular special one-city-pair types, are probably the most commonly encountered model type. An example of a special case trip distribution model is the one in the econometric section in which New York/Florida air traffic is forecast. It is in actuality a direct demand model because it implicitly includes land use, trip generation, and modal split as well as distribution. Model split models predict what percentage of a specific travel market will choose a given mode or carrier.

This chapter began with a discussion of the very general issues regarding model evaluation for all categories. Important issues were then discussed for finer subdivision of the more important category econometric models. We now discuss an even finer set of issues namely those involving statistical evaluation.

A number of econometric models are not statistically valid due to the existence of three fairly common problems: multicollinearity, autocorrelation and heteroscedasticity. Multicollinearity occurs when the explanatory variables are correlated. Its existence tends to produce unreliable coefficients. In addition, they may also be highly volatile. Thus we cannot rely completely on the t-tests. Analysis of simple and partial correlation coefficients can usually give an indication as to the extent and location of intercorrelation among the so called "independent variables."

The major impact of autocorrelation on regression analysis is to cause the calculated error measures to be unreliably estimated. Goodness of fit statistics such as R^2 (coefficient of multiple determination) may have more signficant values than may be warranted. Thus we cannot rely on significant findings obtained from such values to support conclusions about the importance of our relationship. Simple variances of individual regression coefficients may unreliably estimate the true values, causing t-tests to produce possibly spurious conclusions. Finally, the existence of autocorrelation may adversely effect the property of least-squares technique to produce the best estimates. The standard test for this is the Durbin-Watson test. However, if the model is lagged, this test may cause the wrong conclusion to be reached.

Heteroscedasticity occurs when the variance of the error term associated with the fit of an equation is not constant in size across observations. Its existence distorts the measure of unexplained variation, thereby similarly distorting conclusions from R^2 and t-tests based on them. These are a number of tests available to test for heteroscedasticity. They are, however, fairly complicated and require a sound understanding of statistics.

In addition to these three fairly common statistical problems, the following statistical analysis should be performed. First, one should calculate not only the standard error of the forecast but also the standard error of the regression line. The band describing the standard error of the forecast fans out as the regression line is followed further from the range of actual observations. This implies that the confidence which can be placed in a forecast based on trend or regression analysis diminishes very rapidly as the forecast is carried further and further into the future. Second, in testing for the significance of obtained sample R^2 values, an adjustment is necessary to account for statistical bias arising from using sample estimates to measure underlying population characteristics. Third, an evaluation should be made for the impact of omitting significant variables. Fourth, an evaluation should be made by investigating the ANOVA table.

One needs to beware of stability problems in control theory and simulation models. Incorrect specification can result in explosive and unstable models as predictions are made past one's experiences and familiarities. It is difficult to insure a properly behaved model in a range beyond what one can imagine. Computer round-off error can cause problems that one might not expect from just theoretical considerations. Econometric problems can occur from matrices that are nearly non-invertible. Perhaps the most dangerous type of problem is one where large instabilities suddenly occur without warning in otherwise well behaved situations. This problem can occur, for example, when dealing with Kalman filtering (see control theory section). The filtering algorithm makes use of the state covariance matrix (E matrix). Theoretically, the E matrix is symmetric and positive semi-definite. Computer round-off can force the E matrix non-symmetric and non-positive semi-definite. When this happens, instability can occur, and the forecast can very quickly diverge from reality. This problem could be rectified by using a square root of the state covariance matrix, adding non-linear compensation, rectifying the E matrix at each update, or by increasing computer precision. This is just one example of stability problems. Although rare, forecasters always need to keep the possibility of instability in the back of their minds.

"The first and perhaps most obvious requirement of a model for forecasting the consequences of transportation system changes is that the model should be structural; the model should describe the interrelationships among variables which may change in the future, which are predictable (at least within certain limits), and which influence the demand for or performance of a system.... Only by explaining the causal relationships can the model be used to forecast the effects of future changes."[38] This perhaps summarizes the future trend of air travel demand models.

References

1. McDonnell Douglas. "McDonnell Douglas Asked the Experts Their Opinion of Important Future Air Transportation Developments." September 1970.

2. Martino, Joseph P., "Technological Forecasting for Decision Making." New York, American Elsevier Publishing Co. 1972.

3. Martino. Ibid, page 97.

4. Lanford, H.W., "Technological Forecasting Methodologies-- A Synthesis." American Management Association. 1972.

5. Lanford, Ibid, page 79.

6. Lanford, Ibid, page 81.

7. Ayres, Robert U., "Technological Forecasting and Long Range Planning." McGraw-Hill Book Company. 1969.

8. Jantsch, Erich, "Technological Planning and Social Futures." John Wiley & Sons. New York. 1972.

9. Ayres, Ibid, page 72.

10. Steigmann, A. John, "A Partial Recursive Model of Automobile Demand," Business Economics, September 1973, pp. 28-30.

11. Fagan, J., "Stock Price Prediction--A Survey of Current Research and a Comparative Analysis of Two Predictive Techniques." September 1969 (unpublished).

12. Manski, C., Albright, R., & Ben-Akiva, M., "Multinomial Logit Estimation Program." Cambridge Systematics, Inc. October 1973.

13. Wilson, A. (Quandt, R. ed.), "A Statistical Theory of Spatial Distribution Models." <u>The Demand for Travel</u>: <u>Theory and Measurement</u>, 1970.

14. Peat, Marwick, Mitchell & Co., "A Review of Operational Urban Transportation Models." DOT-TSC-496. April 1973.

15. The influence of population growth can be introduced by expressing socio-economic variables in per capita form. For further reference on the influence of population, see G. Besse and G. Deman, "Air Transport, Its 'Conjecture' and the General 'Conjecture' - Past Experience and Lessons for the Future," paper presented at the Second ITA Symposium, November 24-25, 1966, Paris.

16. This is not strictly true since income elasticity starts off being low for the low income levels, rises to its high value for the middle income group and declines again for the high income groups. The basic reasons for a decline in the elasticity for the top income stratum appears to be that people in this group have generally travelled as much as they desire and that further increments of income do not induce proportional changes in travel habits.

17. Alexander P. Triandafyllides, "Forecast of the Demand for North Atlantic Travel," Ph.D. Thesis, University of Washington, 1964; and N.J. Asher et al., Demand Analysis for Air Travel by Supersonic Transport, Institute for Defense Analyses, Washington, D.C., Economic and Political Studies Division; prepared for the Federal Aviation Administration, Contract No. FA-SS-66-14, Arlington, Virginia: 1966.

18. Port of New York Authority, Aviation Economics Division, New York's Overseas Air Passenger Market: April 1963 through March 1964, June 1965; New York's Transatlantic Air Passenger Market: May 1966 through April 1967, July 1969; and New York's Transatlantic Air Passenger Market: May 1968 through April 1969, September 1970.

19. National Planning Association, International Travel on Potential SST Routes, a report on a study performed under U.S. Department of Transportation Contract DOT-FA-SS-71-3, August 1971.

20. Triandafyllides, op.cit.; and U.S. Department of Commerce, Bureau of Economic Analysis, "International Travel, Passenger Fares and Other Transportation in the U.S. Balance of Payments: 1971," reprinted from Survey of Current Business, July 1972.

21. Boeing Company, North Atlantic Macro Air Passenger and Cargo Forecast: An Econometric Approach to Measure Future Demand Levels, November 1971.

22. Philip Verleger presents a good discussion of this point in his article, "Models of the Demand for Air Transportation" in The Bell Journal of Economics and Management Service, autumn 1972, Vol. 3, No. 2.

23. Those interested in weighting the two fares should refer to a good discussion presented in Economics and Tourism: A Study of Factors Affecting Pleasure Travel to the U.S.A. report prepared by Arthur D. Little, Inc. for the United States Travel Service, July 1967.

24. International Air Transport Association, World Air Transport Statistics 1971, Montreal, Quebec: 1972.

25. International Civil Aviation Organization, Digest of Statistics: Traffic Flow - March 1969, June 1969, September 1969, and December 1969, Nos. 148, 149, 150 and 152; Series Nos. 45-48, Montreal, Quebec. Data averaged from the four months.

26. Marc Nerlove, "Distributed Lags and Estimation of Long-Run Supply and Demand Elasticities: Theoretical Expectations," Journal of Farm Economics, Vol XL, No. 2, May 1958.

27. Wayne Watkins and Donna Kaylor, Forecast of Scheduled International Air Traffic of U.S. Flag Carriers 1971-1980, U.S. Civil Aeronautics Board, September 1971.

28. Taneja, N.K. (Howard, G. ed.), "Forecasting Air Passenger Traffic on the North Atlantic," Airport Economic Planning, M.I.T. Press, 1974.

29. It is not feasible to discuss the details of simultaneous equation models or their statistical problems due to the introductory nature of the material presented in this report. The interested reader is referred to the many excellent texts available on the subject, including: Henri Thiel, Principles of Econometrics, John Wiley and Sons, Inc., New York: 1971; and Porituri M. Rao and Roger L. Miller, Applied Econometrics, Wadsworth Publishing Company, Inc., Belmont, California: 1971.

30. Watkins and Kaylor, op.cit.

31. For details of the Chow test see F.M. Fisher, "Test of Equality Between Sets of Coefficients in Two Linear Regressions: An Expository Note," Econometrics, Vol. 38, No. 2, March 1970.

32. National Planning Association, op.cit.

33. Garvett, D., "A Behavioral Approach to Modal Split and Related Problems," December 1973 (unpublished).

34. Ben-Akiva, M., "Structure of Passenger Travel Demand Models," Ph.D. Thesis, M.I.T., Department of Civil Engineering, June 1973.

35. Kraft, G. & Wohl, M., "New Directions for Passenger Demand Analysis and Forecasting," Transportation Research, Vol. 1, No. 3, pp. 205-230, November 1967.

36. Ben-Akiva, M., Ibid.

37. Juster, F., "An Evaluation of the Recent Record in Short-Term Forecasting," Business Economics, May 1972.

38. Kraft, G. & Wohl, M., Ibid.

39. Department of Transportation, "Statistical Analysis of the New York - Washington, D.C. Rail Passenger Service, 1970," October 1971.

Appendix A

Exponentially Weighted Moving Average Program
Non-Adaptive Technique

SECTION 1

Section 1 initializes the program and monitors the search routine as values of α, β and γ are tested to see which combination produces the lowest total squared error in the forecast values.

Section 1.1:	initializes variables, reads in data, does preliminary calculations
Section 1.2:	monitors the search for the best combination of α, β and γ to the nearer .05.
Section 1.3:	monitors the local search around these values to see if a better combination of α, β and γ may be located.

SECTION 2

Section 2 is a subroutine which calculates the S, R and F values for each combination of α, β and γ tested. In addition, it stores the value of the lowest squared error and all the y-values associated with it. If, after a test, a new low total squared error has been found, this value and its associated y-values are stored and the previous low and y-values associated with it are discarded. Then control returns to the main program.

Section 2.1:	finds values for S_1, R_1 and F_1.
Section 2.2:	carries the iterations through the first cycle.
Section 2.3:	carries the iterations through the second cycle.
Section 2.4:	continues the iterations and initiates forecasting.
Section 2.5:	effects a change of stored y-values and the low total squared error value, if necessary.

SECTION 3

Section 3 prints the table.

Section 3.1:	initializes several variables and prints the header lines for the table.

Section 3.2: prints the first four cycles: the two used for backlog values which involve no forecasting as well as the first two forecast periods.

Section 3.3: prints the one remaining cycle.

SECTION 4

Section 4 prints the graphs.

Section 4.1: initializes several variables, prints several header lines and calculates and prints the scale to be used along the y-axis (trip volume) in the first plot.

Section 4.2: prints the volume vs. time graph: a "+" for the true value, a "-" for the forecast value, and a "*" if the two coincide.

Section 4.3: prints the error probability distribution graph.

EXAMPLE

Monthly data showing the total number of trips made by air on scheduled carriers across the North Atlantic for the five years 1967-1971 inclusive is shown in Figure A.1. This can be used to illustrate how the forecasting model operates. Figure A.2 contains a summary of the necessary input data as described above. The output is shown in Figure A.3.

Program Listing of Exponential Smoothing Model

```
$JOB
      SUBROUTINE SECTWO
      COMMON F(59), X(60), Y(60), SAVEY(63), TOTAL, SAVTO
      COMMON ALPHA, BETA, GAMMA, SOLD, SNEW, ROLD, RNEW
      COMMON PINIT, SLOPE, IPER, ITOT, L, M, MA, LA
C
C                             SECTION 2.1
C
      SOLD = (ALPHA * X(1)/F(1)) + ((1.-ALPHA) * (X(I) + SLOPE))
      ROLD = BETA * (SOLD - X(1)) + (1.-BETA) * SLOPE
      F(1) = GAMMA * X(1)/SOLD + (1.-GAMMA) * F(1)
C
C                             SECTION 2.2
C
      DO 4 I = 2, IPER
      SNEW = (ALPHA * X(I)/F(I))+ ((1.-ALPHA) * (SOLD + ROLD))
      RNEW = BETA * (SNEW - SOLD) + (1.-BETA) * ROLD
      F(I) = GAMMA * X(I)/SNEW + (1.-GAMMA) * F(I)
      ROLD = RNEW
    4 SOLD = SNEW
C
C                             SECTION 2.3
C
      DO 5 I = L, M
      SNEW = ALPHA * X(I)/F(I-IPER) + (1.-ALPHA) * (SOLD + ROLD)
      RNEW = BETA * (SNEW - SOLD) + (1.-BETA) * ROLD
      F(I) = GAMMA * X(I)/SNEW + (1.-GAMMA) * F(I-IPER)
      ROLD = RNEW
    5 SOLD = SNEW
C
C                             SECTION 2.4
C
      DO 6 I = LA, MA
      Y(I) = (SOLD + ROLD) * F(I-IPER)
      SNEW = ALPHA * X(I)/F(I-IPER) + (1.-ALPHA) * (SOLD + ROLD)
      RNEW = BETA * (SNEW - SOLD) + (1.-BETA) * ROLD
      F(I) = GAMMA * X(I)/SNEW + (1.-GAMMA) * F(I-IPER)
      ROLD = RNEW
      SOLD = SNEW
    6 TOTAL = TOTAL + ((X(I) -  Y(I))**2)
      Y(I) = (SOLD + ROLD) * F(I-IPER)
      TOTAL = TOTAL + ((X(ITOT) - Y(ITOT))**2)
```

```
C
C                                         SECTION 2.5
C
      IF (SAVTO) 30, 25, 30
   30 IF (TOTAL - SAVTO) 25, 27, 27
   25 DO 26 I = 1, ITOT
   26 SAVEY(I) = Y(I)
      SAVEY(61) = ALPHA
      SAVEY(62) = BETA
      SAVEY(63) = GAMMA
      SAVTO = TOTAL
   27 RETURN
      END
C
      COMMON F(59), X(60), Y(60), SAVEY(63), TOTAL, SAVTO
      COMMON ALPHA, BETA, GAMMA, SOLD, SNEW, ROLD, RNEW
      COMMON PINIT, SLOPE, IPER, ITOT, L, M, MA, LA
      DIMENSION IER(21)
      CHARACTER*80 NAME
      CHARACTER*1 C(36)
      CHARACTER*1 A(130)
      CHARACTER*1 BLANK /'  '/
      CHARACTER*1 DASH /'-'/
      CHARACTER*1 PLUS /'+'/
      CHARACTER*1 STAR /'*'/
C
C                                         SECTION 1.1
C
      ITOT = 60
      IPER = 12
      READ (5, 103) NAME
      READ (5, 101) X
      SAVTO = 0.
      VALAR = 0.
      DO 17 I = 1, 60
      IF (X(I) - VALAR) 17, 17, 18
   18 VALAR = X(I)
   17 CONTINUE
      READ (5, 102) PINIT, PINFI, IYEAR, IUNIT
      PT1 = PINIT
      SLOPE = (PINFI - PINIT)/23.
      DO 59 I = 1, IPER
   59 F(I) = X(I)/(PINIT + ((I - 1) * SLOPE))
      L = IPER + 1
      M = 2 * IPER
      LA = M + 1
      MA = ITOT - 1
```

```
      LB = LA + M
      DO 1 I = 1, M
    1 Y(I) = X(I)
C
C                                         SECTION 1.2
C
      DO 24 IA = 1, 19
      ALPHA = IA * .05
      DO 29 IB = 1, 19
      BETA = IB * .05
      DO 28 IC = 1, 19
      GAMMA = IC * .05
      PINIT = PT1
      TOTAL = 0.
      CALL SECTOW
   28 CONTINUE
   29 CONTINUE
   24 CONTINUE
C
C                                         SECTION 1.3
C
      IA = SAVEY(61)/.05
      IB = SAVEY(62)/.05
      IC = SAVEY(63)/.05
      DO 34 ID = 1, 5
      IF (ID - 3) 31, 34, 31
   31 ALPHA = (IA * .05) + (ID * .01) - .03
      DO 35 IE = 1, 5
      IF (IE - 3) 32, 35, 32
   32 BETA = (IB * .05) + (IE * .01) - .03
      DO 36 IH = 1, 5
      IF (IH - 3) 33, 36, 33
   33 GAMMA = (IC * .05) + (IH * .01) - .03
      PINIT = PT1
      TOTAL = 0.
      CALL SECTWO
   36 CONTINUE
   35 CONTINUE
   34 CONTINUE
C
C                                         SECTION 3.1
C
      DO 49 I = 1, 21
   49 IER(I) = 0.
      ALPHA = SAVEY(61)
      BETA = SAVEY(62)
      GAMMA = SAVEY(63)
```

```
      WRITE (6, 201)
      WRITE (6, 214) NAME
      WRITE (6, 205) ALPHA, BETA, GAMMA, IPER
      WRITE (6, 215)
      WRITE (6, 208)
      TOTAL = 0.
      N = 1
C
C                                    SECTION 3.2
C
      JYEAR  = IYEAR + 2
      DO 22 I = 1, M
      E = X(I+M) - Y(I+M)
      PE = E * 100./X(I+M)
      J = (PE + 2.5)/5
      K = 11 + J
      IER(K) = IER(K) + 1
      EE = E**2
      TOTAL = TOTAL + EE
      WRITE (6, 204) N, IYEAR, X(I), N, JYEAR, X(I+M), SAVEY(I+M), E,
     - PE, EE
      IF (N - 12) 22, 2, 22
    2 IYEAR = IYEAR + 1
      JYEAR  = JYEAR + 1
      N = 0
   22 N = N + 1
C
C                                    SECTION 3.3
C
      DO 23 I = LB, ITOT
      E = X(I) - Y(I)
      PE = E * 100./X(I)
      J = (PE + 2.5)/5
      K = 11 + J
      IER(K) = IER(K) + 1
      EE = E**2
      TOTAL = TOTAL + EE
      WRITE (6, 206) N, JYEAR, X(I), SAVEY(I), E, PE, EE
      IF (N - 12) 23, 3, 23
    3 JYEAR = JYEAR + 1
      N = 0
   23 N = N + 1
      WRITE (6, 207) TOTAL
C
C                                    SECTION 4.1
C
      DO 9 I = 1, 130
```

```
    9  A(I) = BLANK
       B = 10.
   37  IF (B - .01) 42  55, 42
   42  CONTINUE
       DO 19 I = 1, 10
       IF ((VALAR/(I * B)) - 130) 21, 19, 19
   19  CONTINUE
   55  WRITE (6, 216)
       GO TO 39
   21  IF (I - 1) 20, 38, 20
   38  B = B/10.
       GO TO 37
   20  K = I * 20 * B
       KTWO = 2 * K
       KTHR = 3 * K
       KFOR = 4 * K
       KFIV = 5 * K
       KSIX = 6 * K
       WRITE (6, 209)
       WRITE (6, 201)
       WRITE (6, 214) NAME
       WRITE (6, 205) ALPHA, BETA, GAMMA, IPER
       WRITE (6, 202)
       WRITE (6, 213) K, KTWO, KTHR, KFOR, KFIV, KSIX, IUNIT
       WRITE (6, 210)
C
C                                SECTION 4.2
C
       N = 1
       DO 8 J = 1, ITOT
       IORD = X(J)/(I * B)
       JORD = Y(J)/(I * B)
       IF (IORD-JORD) 13, 14, 13
   13  A(IORD) = PLUS
       A(JORD) = DASH
       GO TO 15
   14  A(IORD) = STAR
   15  IF (J/N - 12) 11, 12, 11
   12  WRITE (6, 212) N, A
       N = N + 1
       GO TO 16
   11  WRITE (6, 211) A
   16  A(JORD) = BLANK
    8  A(IORD) = BLANK
       WRITE (6, 203)
```

```
C
C                                    SECTION 4.3
C
 39   WRITE (6, 217)
      K = 0
      N = 5
      DO 41 I = 1, 36
 41   C(I) = BLANK
      DO 48 I = 1, 21
      IORD = IER(1)
      IF (I - 11) 46, 47, 46
 47   IF (IORD) 57, 56, 57
 57   C(IORD) = PLUS
 58   WRITE (6, 219) C
      WRITE (6,218)
      N = 1
      K = -1
      GO TO 56
 46   IF (IORD) 40, 44, 40
 40   C(IORD) = PLUS
 44   IF (N - 5) 50, 51, 50
 51   IF (K**2) 54, 52, 54
 52   J = 50
      K = K + 1
      GO TO 53
 54   J = 25
      K = K + 1
 53   WRITE (6, 220) J, C
      N = 1
      GO TO 56
 50   WRITE (6, 219) C
      N = N + 1
 56   IF (IORD) 43, 48, 43
 43   C(IORD) = BLANK
 48   CONTINUE
      WRITE (6, 209)
      GO TO 10
C
C                                    FORMAT STATEMENTS
C
 101  FORMAT (12(F5.1, 1X))
 102  FORMAT (12X, F5.0, 13X, F5.0, 2(6X, I4))
 103  FORMAT (A80)
 201  FORMAT(' FLIGHT TRANSPORTATION LAB DEMAND FORECASTING MODEL')
 202  FORMAT ('0', 126X, 'VOLUME')
 203  FORMAT (2X, 'YEARS')
 204  FORMAT (3X, I3, '/', I4, 10X, F8.2, 20X, I3, '/', I4, 10X, F8.2,
     -  3(3X, F9.2), 1X, F10.2)
```

```
205  FORMAT(' ALPHA = ', F4.2, ', BETA = ', F4.2, ', GAMMA = ', F4.2,
    - ', PERIOD = ', I2)
206  FORMAT (49X, I3, '/', I4, 10X, F8.2, 3(3X, F9.2), 1X, F10.2)
207  FORMAT (110X, F12.2, '***')
208  FORMAT (23X, 'VOLUME', 40X, 'VOLUME', 7X, 'VOLUME', 17X, 'ERROR',
    - 7X, 'ERROR')
209  FORMAT ('1')
210  FORMAT ('  +-------------------|-------------------|------------',
    -   '-------|-------------------|-------------------|-------------',
    -   '------|----------')
211  FORMAT (2X, '|', 130(A1))
212  FORMAT (1X, I1, '-', 130(A1))
213  FORMAT '   0', 6(16X, I4), 4X, 'X ', I4)
214  FORMAT (1X, A80)
215  FORMAT ('0MONTH/YEAR', 12X, 'ACTUAL', 18X, 'MONTH/YEAR', 12X,
    -   'ACTUAL', 6X, 'FORECAST', 5X, 'ERROR', 5X, 'PERCENT', 5X
    -   'SQUARED')
216  FORMAT ('1VOLUMES TOO LARGE TO PERMIT GRAPHICAL DISPLAY')
217  FORMAT ('0ERROR PROBABILITY DISTRIBUTION')
218  FORMAT  ('+ 0%-----------------------------------1')
219  FORMAT (4X,  '|', 36(A1))
220  FORMAT (1X, I2, '%-', 36(A1))
 10  CONTINUE
     STOP
     END
```

NORTH ATLANTIC TRAFFIC
MONTHLY TOTALS
1967-1971

	1967	1968	1969	1970	1971
January	107,597	119,795	131,609	154,349	163,945
February	84,978	102,252	109,753	140,412	142,927
March	120,782	130,890	157,181	207,079	209,614
April	161,809	177,333	192,413	232,054	247,774
May	236,844	222,034	255,032	341,836	364,161
June	313,714	330,950	386,550	457,655	445,223
July	368,250	434,008	507,720	590,425	588,906
August	296,425	292,871	343,100	464,504	403,102
September	257,260	252,369	290,552	348,486	438,999
October	189,551	186,147	206,580	225,548	270,639
November	105,746	121,262	140,428	153,020	179,118
December	138,517	159,581	200,349	212,854	238,663

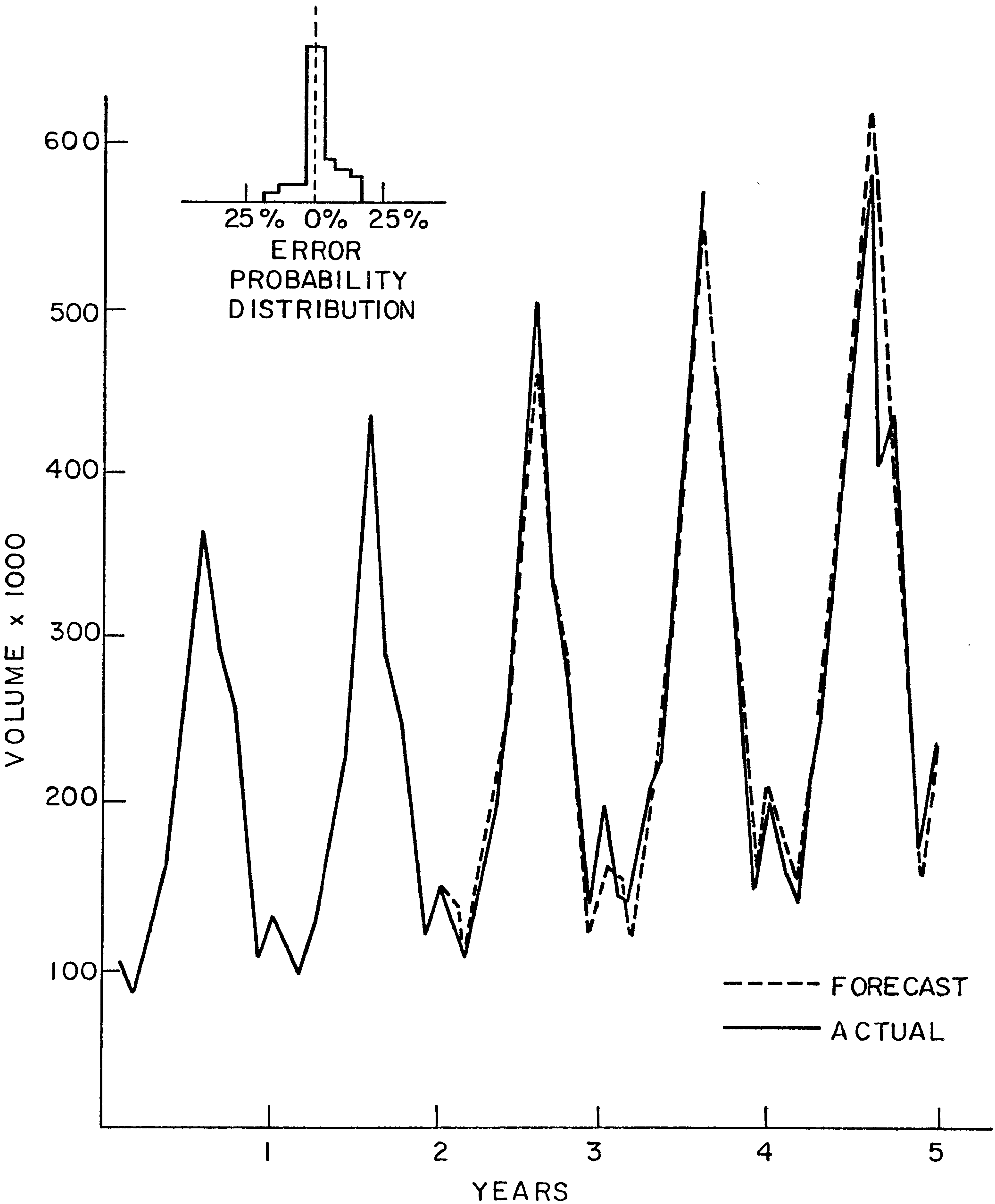
Figure A.1.

NORTH ATLANTIC TRAFFIC FORECAST 1967-1971

107.5	085.0	120.8	161.8	236.8	313.7	368.3	296.4	257.3	189.6	105.7	133.5
119.8	102.3	130.9	177.3	222.0	331.0	434.0	292.9	252.3	185.1	121.3	159.6
131.6	109.8	157.2	192.4	255.0	386.6	507.7	343.1	290.6	206.5	140.4	205.3
154.3	140.4	207.1	232.1	341.8	457.7	590.4	464.5	348.5	225.5	153.0	212.9
163.9	142.9	209.6	247.8	364.2	445.2	588.9	403.1	439.0	270.6	179.1	238.7

FIRST POINT 179 FINAL POINT 240 YEAR 1967 UNIT 3

Figure A.2.



FLIGHT TRANSPORTATION LABORATORY DEMAND FORECASTING MODEL NORTH ATLANTIC TRAFFIC FORECAST 1967–1971
alpha = 0.15 beta = 0.05 gamma = 0.40 PERIOD = 12

Figure A.3

Appendix B

Spectral Analysis

This program calculates the mean, variance, and auto-correlation function of a time series. It also prints out each frequency considered and the power (spectral density) associated with it. At the beginning of the output, the input data is reproduced for future reference. One then has available the underlying frequencies of the series and their amplitudes and can then evaluate the series.

The program is written in Fortran and uses namelist input.* The user has to supply the number of observations in the series, the number of lags he wants to include in the analysis, and the series itself. The program can process up to five hundred observations and one hundred lags. Changing the dimension statements can increase these limits. The series is assumed to be stationary. Proper logarithmic transformations or generalized differencing should be done if needed to insure stationarity.

```
C ************************************************
C SPECTRAL ANALYSIS PROGRAM
C ************************************************
C THIS PROGRAM ASSUMES THAT THE SERIES THAT IS SUPPLIED IS STATIONARY
      REAL MEAN
      DIMENSION VALUE(500),AUTOCO(100),PDS(200),TEMP(100)
C PDS IS THE POWER DENSITY SPECTRUM
      NAMELIST/INPUT/NOBS,LAGS,VALUE
      DATA PI/3.14159265/
```

*If the user is unfamiliar with the required job control language or namelist input, he should consult an IBM Fortran manual.

```
C
 1    FORMAT ('1 MEAN=',E15.7,' VARIANCE=',E15.7/)
 2    FORMAT (' LAG',6X,' AUTO-',5X,' FREQUENCY',2X,' SPECTRAL'/,7X,
     1 ' CORRELATION',13X,' DENSITY')
 3    FORMAT (I4,F11.4,F12.4,F10.4)
 4    FORMAT (15X,F12.4,F10.4)
C
      READ (5,INPUT)
      WRITE (6,INPUT)
      LAGS = LAGS + 1
      MEAN = 0.
      DO 10 I=1,NOBS
10    MEAN = MEAN + VALUE(I)
      MEAN = MEAN / FLOAT(NOBS)
      LL = 1 + LAGS / 2
      LL1 = LL + 1
      L1 = LAGS - 1
      FL1 = L.
      L2 = LAGS * 2
      ANGLE = PI / (FL1 * 2.)
      DO 25 I=1,NOBS
25    VALUE(I) = VALUE(I) - MEAN
      DO 20 J=1,LAGS
      AUTOCO(J) = 0.
      NJ = NOBS - J + 1
      DO 20 I=1,NJ
      IJ = I + J - 1
20    AUTOCO(J) = AUTOCO(J) + VALUE(I) * VALUE(IJ)
      DO 30 J=2,LAGS
30    AUTOCO(J) = AUTOCO(J) / AUTOCO(1)
      VAR = AUTOCO(1) / FLOAT(NOBS)
```

```
      AUTOCO(1) = 1.
      DO 4- J=2,LL
      TAU = J - 1
40    TEMP(J) = 1. - (6. * TAU**2 / (FL1**2)) + (6. * TAU**3/(FL1**3))
      DO 50 J=LL1,L1
      TAU = J - 1
50    TEMP(J) = 2. * (1. - TAU / FL1)**3
      DO 60 I=1,L2
      PDS(I) = 0.
      F11 = I-1
      DO 60 J=2,L1
      FJ1 = J - 1
60    PDS(I) = PDS(I) + AUTOCO(J)*TEMP(J)*COS(ANGLE*FJ1*FI1)
      WRITE (6,1) MEAN,VAR
      WRITE (6,2)
      DO 70 I=1,LAGS
      I1 = I - 1
      FI1 = I1
      PDS(I) = (1. + 2. * PDS(I)) * 2
      FREQ = FI1 / (4. * FL1)
70    WRITE (6,3) I1,AUTOCO(I), FREQ PDS(I)
      L1 = LAGS + 1
      L2 = L2 - 1
      DO 80 I=L1,L2
      PDS(I) = (1. + 2. * PDS(I)) * 2.
      FI1 = I - 1
      FREQ = FI1 / (4. * FL1)
80    WRITE (6,4) FREQ,PDS(I)
      STOP
      END
```

Appendix C

Florida and Orlando Air Travel Models

Florida

1. $Volfla = a*Volfla(-4)^{b}*popq^{c}*Edisney^{d}*Farefla^{e}*Income^{f}*Eweather^{g}$
2. $Volfla = a*Volfla(-4)^{b}*Edisney^{d}*Farefla^{e}*Income^{f}*Eweather^{g}$
3. $Volfla = a*Volfla(-4)^{b}*Edisney^{d}*Farefla^{e}*Income^{f}$
4. $Volfla = Volfla(-4)^{b}*Edisney^{d}*Farefla^{e}*Income^{f}$
5. $Volfla = Volfla(-4)^{b}*Farefla^{e}*Income^{f}$
6. $Volfla = a*Volfla(-4)^{b}*Edisney^{d}*Income^{f}*Eweather^{g}$
7. $Volfla = a*Volfla(-4)^{b}*Edisney^{d}*Income^{f}$
8. $Volfla = Volfla(-4)^{b}*Edisney^{d}*Income^{f}$
9. $Volfla = a*Edisney^{d}*Farefla^{e}*Eweather^{g}$
10. Volfla = a + d*Disney + e*Farefla + f*Income + g*Weather
11. Volfla = a + b*Volfla(-4) + d*Disney + e*Farefla + f*Income + g*Weather
12. Volfla = a + d*Disney + e*Farefla + f*Income + h*Winter
13. Volfla = d*Disney + e*Farefla + f*Income + h*Winter
14. Volfla = a + b*Bolfla(-4) + d*Disney + e*Farefla + f*Income
15. Volfla = b*Volfla(-4) + d*Disney + e*Farfla + f*Income

Orlando

1. Volorl = a + b*Volorl(-4) + d*Disney + e*Fareorl + f*Income + g*Weather
2. Volorl = a + d*Disney + e*Fareorl + f*Income + g*Weather
3. Volorl = a + b*Volorl(-4) + d*Disney + e*Fareorl + f*Income
4. Volorl = a + b*Volorl(-4) + d*Disney + f*Income
5. Volorl = a + d*Disney + f*Fareorl
6. Volorl = a + d*Disney + e*Fareorl + f*Income + h*Winter
7. Volorl = d*Disney + e*Fareorl + f*Income + h*Winter
8. Volorl = a + d*Disney + f*Income

Appendix D

Description of TUBSIM

The input for this TUBSIM model is in the form of a namelist*. Required input consists of the reject criteria (mean and standard deviation for each market segment and each level of service characteristics), the level of service values (mean and standard deviation for each market segment, each level of service characteristic, and each mode), the values of the parameters of the utility function (mean and standard deviation for each market segment and each level of service characteristic), the number of passengers to consider in the simulation, the auto availability of the lower income groups (upper income groups assumed to have complete auto availability), and the fraction of the total market that each segment makes up. The reject criteria, level of service characteristics, and auto availability are determined from market surveys. The fraction of the market for each segment is determined from an exogenous model or other source. The parameters of the utility function are determined by logit models. The logit models' output includes the estimated value of the parameters, the standard errors of the estimates, the t-statistics, and a pseudo-R-squared. The number of passengers is increased until stable results are obtained. One should start with approximately one thousand passengers.

The output consists of the absolute number and the percentage of each market segment that each mode captures. An aggregate modal split is also calculated. The 'no mode' alternative shows how many passengers rejected all modes. The high rejection rate (5-6%) in these tests are due to the small number of market segments and minor statistical problems which can easily be corrected when the

*If unfamiliar with the namelist input, one should consult an IBM Fortran manual.

technique is applied to serious forecasting.

TUBSIM Program Listing

```
      SUBROUTINE SEG(N,IX)
      COMMON/TRAVEL/SHARE
      DIMENSION SHARE (6)
C SHARE SPECIFIES WHAT FRACTION OF THE TOTAL MARKET
C EACH MARKET SHARE MAKES UP
C ***********************************************
C 1 SINGLE-DAY BUSINESS TRIP
C 2 MULTI-DAY BUSINESS TRIP
C 3 SINGLE-DAY PERSONAL TRIP (>$10,000)
C 4 SINGLE-DAY PERSONAL TRIP (<$10,000)
C 5 MULTI-DAY PERSONAL TRIP (>$10,000)
C 6 MULTI-DAY PERSONAL TRIP (<$10,000)
C ***********************************************
      CALL RANDU(IX,IY,RAND)
      IX=IY
      SUM = SHARE(1)
      IF (RAND.GT.SUM) GO TO 10
      N=1
      RETURN
10    SUM = SUM + SHARE(2)
      IF (RAND.GT.SUM) GO TO 20
      N=2
      RETURN
20    SUM = SUM + SHARE(3)
      IF (RAND.GT.SUM) GO TO 30
      N=3
      RETURN
```

```
30    SUM = SUM + SHARE(4)
      IF (RAND.GT.SUM) GO TO 40
      N=4
      RETURN
40    SUM = SUM + SHARE(5)
      IF (RAND.GT.SUM) GO TO 50
      N=5
      RETURN
50    N=6
      RETURN
      END

      SUBROUTINE GENLOS(N,VALUE,IX,STATS)
      DIMENSION VALUE(4,4),STATS(6,4,4,2)
C VALUE CONTAINS THE VALUES FOR PAX N FOR TIME,COST
C SKED DELAY, AND CC IN THAT ORDER FOR LOS
C
C STATS PARAMETERS=SEGMENT,MODE,LOS MEAN OR SIGMA
      DO 10 I=1,4
      DO 10 J=1,4
      CALL GAUSS(IX,STATS(N,I,J,2),STATS(N,I,J,1),RANDOM)
10 VALUE(I,J) = AMAX1(RANDOM,0.)
      RETURN
      END
      SUBROUTINE VALUES(N,VALUE,IX,STATS)
      DIMENSION VALUE(4),STATS(6,4,2)
C STATS INCLUDES (1) MEANS and (2) SIGMA
C FOR EACH GROUP AND REJECT CRITERIA OR DISUT
      DO 10 I=1,4
      CALL GAUSS(IX,STATS(N,1,2),STATS(N,I,1),RANDOM)
10    VALUE(I) = AMAX1(RANDOM,0.)
      RETURN
      END
```

```
C TUBSIM
      REAL IMPEDE, LOS, LOSN
      INTEGER PAX
               COMMON/TRAVEL/SHARE
C CODE FOR IREJEK & PAX 1=AIR,2=BUS,3=AUTO,4=RAIL,5=NONE
   DIMENSION REJECT(6,4,2),LOS(6,4,4,2),DISUT(6.4.2),IREJEK(4),

     1LOSN(4,4),DISUT(4),REJEKN(4),PAX(7,5),SHARE(6)
      NAMELIST/TUBSIM/REJECT,LOS,DISUT,NPAX,AUTO,SHARE
C AUTO IS THE FRACTION OF AUTO AVAIL FOR SEG 4 & 6; FOR
C OTHER GROUPS, 100% AUTO AVAILABILITY IS ASSUMED
      READ(5,TUBSIM)
      WRITE(6,TUBSIM)
      IX = 1
      DO 10 I=1,7
      DO 10 J=1,5
      PAX(I,J) = 0
10 CONTINUE
110 FORMAT(2X,I4,2X,I4,2X,I4,2X,I4,2X,I4)
120 FORMAT(//'MODAL SPLIT (ABSOLUTE) FOR SEGMENT',I3)
130 FORMAT('MODAL SPLIT (PERCENT) FOR SEGMENT',I3)
131 FORMAT('1 AIR BUS AUTO RAIL NO MODE')
      DO 999 II=1,NPAX
      DO 20 I=1,4
20   IREJEK(I) = 0
      CALL SEG(N,IX)
      IF (N .NE. 4 .AND. N .NE. 6) GO TO 30
      CALL RANDU(IX,IY,RAND)
      IX = IY
      IF (AUTO .LT. RAND) IREJECK(3) = 1
```

```
30    CONTINUE
       CALL GENLOS(N,LOSN,IX,LOS,BAND)
       CALL VALUES(N,DISUTN,IX,DISUT,BAND)
       CALL VALUES(N,REJEKN,IX,REJECT,BAND)
       DO 160 L=1,4
160   LOSN(L,4) = AMIN1(LOSN(L,4),5.)
       DO 40 I=1,4
C ABOVE CHOOSES MODE
       DO 50 J=1,4
C ABOVE CHOOSES PARAMETER
       IF (LOSN(I,J) .LT. REJEKN(J))GO TO 50
       IREJEK(I) = 1
       GO TO 40
50    CONTINUE
40    CONTINUE
       MODE = 5
       BEST = 1.E60
       DO 60 I=1,4
C ABOVE CHOOSES MODE
       IF (IREJEK(I) .EQ. 1) GO TO 60
       IMPEDE = 0.
       DO 70 J=1,4
70    IMPEDE = IMPEDE + DISUTN(J) * LOSN(I,J)
       IF (IMPEDE .GT. BEST) GO TO 60
       BEST = IMPEDE
       MODE = I
60    CONTINUE
       PAX(N,MODE) + 1
999   CONTINUE
       DO 90 I=1,5
       DO 90 J=1,6
```

```
90    PAX(7,1) = PAX(7,1) + PAX(J,I)
      WRITE(6,131)
      DO 80 I=1,7
      WRITE(6,120)I
      WRITE(6,110)(PAX(I,J),J=1,5)
      IDUM = PAX(I,1)
      DO 140 J=2,5
140   IDUM = IDUM + PAX(I,J)
      DO 150 J=1,5
150   PAX(I,J) = (100. * FLOAT(PAX(I,J))/FLOAT(IDUM) + .5)
      WRITE(6,130)I
      WRITE(6,110)(PAX(I,J),J=1,5)
80    CONTINUE
      STOP
      END
```