Time-Series Analysis of Intercity Air Travel Volume

PHILIP J. OBERHAUSEN AND FRANK S. KOPPELMAN

This research develops a useful model from which to analyze intercity air travel demand and to produce short-term forecasts. Traditional techniques are presented and technical issues associated with these techniques are discussed. An alternative procedure developed by Box and Jenkins is then introduced. This procedure can be used to develop univariate models that account for monthly as well as seasonal patterns in a time series of historical data. Explanatory variables may also be added to form multivariate models. The technique involves four stages: identification, estimation, diagnostic checking, and forecasting. The Box and Jenkins methodology is applied to a monthly time series of visitor air travel from mainland North America to Hawaii. A univariate model is developed with monthly data from 1971 through 1978, and variations of the model are statistically compared. Forecasts based on the "best" univariate model are then computed for 1979 and 1980 and compared with actual data. Results show that the univariate model selected produced reasonably accurate short-term forecasts. Some 17 of 23 forecasts are not significantly different from the actual observations. When updated, these forecasts are even more accurate. Finally, a bivariate time-series model incorporating air fare as an explanatory variable is estimated. It does not produce a significantly better fit of the data in this case. However, these models are potentially useful from a management standpoint because elasticities can be derived and alternative strategies analyzed. In the Hawaii air travel market, additional research is needed to refine the underlying variable relations and their influence on demand.

The commercial air transportation industry has experienced tremendous growth since the middle of this century. However, current and potential carriers are faced with decisions in the 1980s that will determine their future prosperity if not survival. Recent developments in the industry, including deregulation and increasing fuel prices, are forcing carriers to make critical decisions with regard to fare pricing, fleet expansion, route structure, and flight scheduling. In the public sector, air terminal authorities are faced with serious problems resulting from the rapid growth of commercial and private air transportation in their communities. From these perspectives, decisionmakers need to understand the dynamics of the public demand for air transportation and, it is hoped, how their decisions interact with that demand.

This research is concerned with the analysis and forecasting of intercity air travel demand. The particular market chosen for study is that of visitor travel from mainland North America to the Hawaiian Islands. The importance of such a study goes beyond the frame of reference of air carrier or airport management. The notion of transportation as a derived demand is particularly clear in this market, where a vacation in Hawaii is the dominant trip purpose. From this perspective, travel demand patterns are also of major concern to those involved with the entire Hawaii visitor industry, including hotel, entertainment, and other service establishments.

AIR TRAVEL FORECASTING BACKGROUND

There are several ways to categorize air travel forecasting methods. One of the more general distinctions is between purely judgmental approaches and mathematical modeling.

Judgmental methods elicit the personal opinions and predictions of experts in the various fields of air transportation. A popular technique used to obtain information in this way is the Delphi method, where several experts respond independently to several questions pertaining to future air travel demand (1). After seeing their fellow experts' predictions and reasoning, participants are given the opportunity to change their estimates. The intention is that some consensus will eventually be reached and that this consensus will be a good estimator of future demand. Problems with this method include the determination of consensus criteria and the possibility that responses will polarize rather than come together.

The other general procedure used to predict air travel is based on the use of mathematical models. The five-step procedure used to develop these models for prediction is well-established and it includes the following:

1. Variable specification,
2. Variable measurement,
3. Model formulation,
4. Model estimation, and
5. Policy analysis and forecasting.

One of the simplest types of air travel forecasting models relates the amount of travel observed to time. Models of this type are called trend extrapolation models and only one variable, namely the amount of travel, needs to be measured. An histor-
Figure 1. Box and Jenkins time-series analysis procedure.

- **Identification**
  - NO
  - Model Identification
  - Parameters Acceptable and Significant?
    - NO
    - Diagnosis of Residuals
    - Model Adequate?
      - NO
      - FORECAST
      - YES

1. A nationwide personal income index; and
2. A nationwide unemployment rate.

The FAA procedure incorporates projected geographic shifts in income, which is important, and it develops separate equations for originating, returning, and connecting passengers.

**Box and Jenkins Approach to Time-Series Analysis**

This research develops both a trend model and a temporal-structural model of the demand for air travel from mainland North America to Hawaii. These models attempt to fit the patterns observed in a time series of historical data. The patterns are described by the autocorrelation of observations in a single series such as the observed number of air travelers in a market. Time-series analysis is used to study the autocorrelation patterns between successive observations in the time series and, in some cases, patterns between successive seasonal observations.

In 1976, Box and Jenkins developed a simple procedure for identifying and modeling the autocorrelation patterns within a time series (5-7). This procedure, diagrammed in Figure 1, includes three basic phases: (a) identification of a tentative model, (b) estimation of model parameters, and (c) diagnostic checking of residuals.

**Identification**

Identification of the tentative model is accomplished through observing the patterns of the autocorrelation function (ACF) and the partial autocorrelation function (PACF) of the series in question. The ACF at lag one is a combined measure of correlation between each value in the series and the value one period behind it; the ACF at lag two compares each value and the value two periods behind, etc. An exponentially decreasing pattern observed in the ACF is an indication of a particular type of autocorrelation and the resulting model is referred to as an autoregressive model. The number of autoregressive parameters to include in such a model is given by the number of significant values observed in the PACF. The ACF and PACF also indicate whether the data include a seasonal pattern. Once identified, the tentative model structure can be written. For example, by using monthly data, an autoregressive model with a single component of autocorrelation for successive values and a single seasonal component of autocorrelation for values 12 months apart can be represented as

\[
(1 - \phi_1 B)(1 - \phi_{12} B^{12})y_t = \epsilon_t
\]

where

- \( Y_t \) = series observation at time \( t \)
- \( \phi_1 \) = autoregressive parameter indicating the relationship between successive values in the series
- \( \phi_{12} \) = autoregressive parameter indicating the seasonal relation between values in the series from year to year,
- \( B \) = backshift operator on \( Y \) such that \( B^{12} = Y_{t-12} \)
- \( \epsilon_t \) = residual error between observed and fitted values

Expanding on Equation 2 we obtain

\[
Y_t = \phi_1 Y_{t-1} + \phi_{12} Y_{t-12} - \phi_1 \phi_{12} Y_{t-13} + \epsilon_t
\]

Finally, eliminating the backshift operator notation, we have
Estimation

Once a tentative model has been identified, estimation of the autoregressive parameters is performed. There are several computer routines available to do this (8, 9). Parameter values are checked for significance based on the level desired. Insignificant parameters are an indication of a misspecified or an overspecified model. All autoregressive parameters should fall between -1 and +1 for the model to be acceptable.

Diagnostic Checking

After the parameters have been estimated, the ACF and PACF patterns of the residuals are checked to see if there is any remaining autocorrelation unaccounted for by the model. Significant values at the early or seasonal lags are indicators that the model is underspecified and requires additional parameters. This phase will not detect an overspecification. A goodness-of-fit measure can be computed from the residual ACF and the null hypothesis that the model is adequate can be tested statistically. Once the model is deemed adequate, one may proceed with forecasting.

Model Verification

In an attempt to verify the Box-Jenkins methodology, this research includes an additional step, which compares the model obtained by the Box-Jenkins procedure with three alternative models and uses statistical tests to identify the best model.

The model represented by Equation 4 is basically a sophisticated trend model that contains no explanatory variables. It is referred to as a univariate model since only one time series, that representing the behavior itself, is analyzed. However, the Box-Jenkins philosophy allows for inclusion of explanatory variables in these time-series models. In this study, we first develop a univariate model of air travel from the mainland to Hawaii. This model is underspecified compared with other models and its forecasts are validated by comparison to actual observations. Then we incorporate an explanatory variable that is hypothesized to influence demand. The resulting bivariate time-series model is compared statistically with the univariate model and analyzed from a management policy standpoint.

UNIVARIATE ANALYSIS OF TRAVEL VOLUME

We analyzed westbound visitor travel to Hawaii from mainland North America because this market includes mostly domestic traffic. Those arriving from the Far East and the South Pacific are excluded so as to minimize international factors and facilitate simpler data collection. The Hawaii Visitors Bureau (HVB) compiles monthly information on the number of westbound visitors destined for Hawaii. All passengers on flights bound for Hawaii are asked to fill out a questionnaire about themselves and their current trip. These forms are then tabulated and each year the HVB publishes its annual research report summarizing these statistics.

For our purposes, data were collected from the HVB annual reports for nine years from January 1971 through December 1979. Figures for the first 11 months of 1980 were subsequently obtained through personal correspondence.

Figure 2 is a plot of the time-series of westbound visitor travel to Hawaii. It exhibits an overall increasing trend and a seasonal pattern. The months of March and August are generally the heaviest in any given year, while May and September show the lowest amount of activity.

Several computer packages are available to perform univariate time-series analysis. One of the most convenient is SCRUNCH/SCRTIME, which was developed at Northwestern University and used in this research (8). The analysis is done on the data through December 1978 so that forecasts can be compared with actual observations in 1979 and 1980.

The ACF of the travel volume time series is shown in Figure 3. The exponentially decreasing pattern that is evident in the early lags is an indication of an autoregressive model. The number and types of autoregressive parameters to be included in the model are found by examining Figure 4, the PACF. Significant values at lag one and (marginally) at lag two mean that observations in the time series are related to previous values one and two periods before. The "spike" at lag 12 indicates a seasonal relation between observations 12 months apart.

The parameter estimates for this model are as follows:

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\phi_1$</td>
<td>0.377</td>
<td>0.140 to 0.615</td>
</tr>
<tr>
<td>$\phi_2$</td>
<td>0.333</td>
<td>0.088 to 0.578</td>
</tr>
<tr>
<td>$\phi_{12}$</td>
<td>-0.383</td>
<td>-0.618 to -0.147</td>
</tr>
<tr>
<td>$\phi_0$</td>
<td>6660.0</td>
<td>825.0 to 124.91</td>
</tr>
<tr>
<td>(constant)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Since all are significant at the 95 percent level, we proceed to the third step in the process, diagnosis.

The ACF of the residual series generated from the estimated model is shown in Figure 5. The model appears to be adequate since there are no significant residual autocorrelations through the first 30 lags. The Q-statistic reported by the estimation program is an overall measure of the adequacy of the model. This statistic can be used to test the null hypothesis that our model provides an adequate fit.

\[
Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-12} + \phi_1 Y_{t-13} + \epsilon_t
\]
of the data by comparison with the chi-square distribution. In this case, the Q-statistic is 16.7, which is well within the 95 percent confidence limit of 38.9 for 26 degrees of freedom (DF). This means that we fail to reject the null hypothesis and, for now, conclude that the specified model is an adequate representation of westbound visitor travel to Hawaii from 1971 to 1978.

As an additional step, we attempted to verify the model obtained by using the Box-Jenkins methodology by statistically comparing it with three alternative autoregressive models. Summary estimation results and fit statistics of the four models are given in Table 1. Model 1 is the simplest model, models 2 and 3 are intermediate, and model 4 is the complete model identified above. F-ratios were computed to test whether the models with more parameters were significantly better than the simpler model. Figure 6 is a diagram showing the results of the statistical tests. Model 1 is rejected by all three larger models, and both model 2 and model 3 are rejected by model 4. This verification is encouraging, since it indicates that the model selected by using the Box-Jenkins procedure is the best statistical model as well.

By using Box-Jenkins notation, the model we have selected may be expressed as follows:

\[
(1 - \phi_1 B)(1 - \phi_2 B^2)(1 - \phi_3 B^3)Y_t = \phi_4 + \epsilon_t
\]

(5)

The first term on the left side of Equation 5 is a seasonal differencing factor, required prior to identification due to the upward seasonal trend of the data. Substituting the parameter estimates into Equation 6 and expanding, we obtain

\[
Y_t = 0.38Y_{t-1} + 0.33Y_{t-2} + 0.26Y_{t-3} - 0.23Y_{t-4} + \epsilon_t
\]

- 0.30Y_{t-14} + 0.38Y_{t-24} - 0.14Y_{t-25} - 0.13Y_{t-26} + 660 + \epsilon_t

(6)

Because this equation is designed to fit the time series of data in Figure 2, it is necessarily somewhat complex. However, the important point is that it is relatively easy to identify this model by using the Box-Jenkins procedures.

---

Table 1. Alternative univariate models

<table>
<thead>
<tr>
<th>Seasonal Component</th>
<th>Monthly Component</th>
</tr>
</thead>
<tbody>
<tr>
<td>((1 - \phi_1 B))</td>
<td>((1 - \phi_2 B^2))</td>
</tr>
<tr>
<td><strong>Model 1</strong></td>
<td><strong>Model 2</strong></td>
</tr>
<tr>
<td>Trend = 8474.0</td>
<td>Trend = 5816.6</td>
</tr>
<tr>
<td>(\phi_1 = 0.477 48)</td>
<td>(\phi_1 = 0.343 80)</td>
</tr>
<tr>
<td>(\phi_2 = 0.310 42)</td>
<td>(\phi_2 = 0.333 04)</td>
</tr>
<tr>
<td>RSSQ = 0.11467 x 10^11</td>
<td>RSSQ = 0.10489 x 10^11</td>
</tr>
<tr>
<td>NOBE = 70</td>
<td>NOBE = 70</td>
</tr>
</tbody>
</table>

\[1 - (1 + \phi_1 B^2)(1 + \phi_2 B^2)\]

**Model 3**

| Trend = 10662.0 | Trend = 6635.6 |
| \(\phi_1 = 0.543 62\) | \(\phi_1 = 0.377 51\) |
| \(\phi_2 = -0.366 68\) | \(\phi_2 = 0.333 04\) |
| \(\phi_4 = -0.382 85\) | \(\phi_4 = 0.333 04\) |
| RSSQ = 0.1004 x 10^10 | RSSQ = 0.09023 x 10^10 |
| NOBE = 70 | NOBE = 70 |
| DF = 67 | DF = 66 |

**Note:** RSSQ = residual sum of squares, NOBE = number of effective observations, NFAR = number of model parameters, and DF = degrees of freedom = NOBE - NFAR.
It can be seen from Equation 6 that the value of the series at time “t” is positively related to values 1, 2, 12, and 24 months before. (The negative coefficients for lags of 13, 14, 25, and 26 months are terms that eliminate double counting of successive period and seasonal effects.) Practically, this means that a monthly increase or decrease in travel tends to perpetuate itself successively for 2 months and seasonally for 2 years.

By using the univariate model selected above, forecast ranges were computed at a 95 percent level of confidence for each month from January 1979 through November 1980. These are plotted as the shaded areas in Figure 7.

Comparing these forecast intervals to actual experience, we see that 10 of the 12 months in 1979 were predicted within the 95 percent confidence range. The model significantly overpredicted the months of April and May. This is most probably explained by the fact that United Air Lines, with a large market share (more than 50 percent) of mainland-to-Hawaii air travel, suffered a work stoppage between March 31 and May 28, 1979. In 1980, the months of January, September, October, and November were significantly overpredicted by the model. In fact, most months in 1980 are only narrowly within the lower bound of the prediction range. This is most likely explained by the recent recession that caused the visitor industry in Hawaii to experience a pronounced slowdown in 1980. The univariate model developed through 1979 would not be able to predict this change in trend. This suggests the need to consider inclusion of descriptive variables to account for changing economic conditions.

In practice, of course, forecasts should be updated as additional data points become available. Two sets of updated forecasts for 1979 and 1980 by using the original model parameters were computed, one assuming immediate updating of information and one assuming updating with a 3-month delay. Because the United Air Lines strike represents a particularly sharp anomaly in the data for the months of April and May 1979, the data for this period were adjusted to the predicted volume in the absence of the work stoppage.

The effect of updating the travel demand information prior to forecasting a new month is to produce more accurate forecasts. The root-mean-square error (RMSE) is a measure of the overall deviation of a series of forecasts from actual experience. This measure decreased by 15 percent for updating with a 3-month delay and 32 percent when the series was updated with a 1-month delay prior to forecasting.

MULTIVARIATE ANALYSIS

We are interested in the way in which the travel volume series may be influenced by changes in exogenous variables. We chose to examine the effect of the price of air travel between mainland North America and Hawaii. Round-trip coach fares between Hawaii and four major metropolitan areas in North America were weighted according to the volume of Hawaii travel observed from those areas. A monthly time series of these average coach fares was computed and deflated by the consumer price index for each period from January 1971 to December 1978.

By using the multivariate time-series analysis
computer package called WMTS-1 (10), the average coach fare variable is incorporated and a transfer function model is estimated. The effect of price changes on demand one period later is observed. The same system of autoregressive components identified above is used here. The estimated equation for this bivariate model is as follows:

\[
Y_t = 0.30Y_{t-1} + 0.34Y_{t-3} + 0.62Y_{t-12} - 0.22Y_{t-3} - 0.21Y_{t-14} + 0.38Y_{t-24} - 0.14Y_{t-23} - 0.13Y_{t-26} - 0.07X_{t-1} + 279.92 + \epsilon_t
\]

where \( Y \) is visitor travel from the mainland to Hawaii and \( X \) is the average round-trip coach fare from the mainland to Hawaii.

We notice that the autoregressive parameters for this model are virtually identical to those in Equation 6, as we would expect. The coefficient for average coach fare denotes the effect that a change in fare will have on demand. In this case, the magnitude and significance of the fare coefficient are rather low. Based on these results, the price elasticity of demand is approximately -0.1. In terms of overall fit, the bivariate model is not significantly better than the univariate model, suggesting either that the Hawaii market is rather inelastic with respect to price or that we have not captured the true price effect with this particular measure.

In fact, we believe that a preferred measure of cost should be an estimate of total cost for the Hawaii vacation including air fare, the cost of accommodations, and other local expenses. If air fare is approximately one-third of total vacation cost, the elasticity of demand to total cost implied by the results reported here would be in the range of -0.3.

POLICY ANALYSIS

One of the advantages of multivariate analysis is that policymakers and air carrier management can use the structural parameters to develop forecasts based on alternative future scenarios. This was done by using the bivariate model estimated above. Two alternative air fare pricing policies were analyzed in terms of their effect on forecasted travel for 1979. Air carriers have been faced with conflicting economic pressure to both raise fares due to rising costs of labor and fuel on the one hand, and to lower fares for competitive reasons on the other. Thus, the two alternative pricing strategies studied were (a) a 10 percent quarterly increase in the average coach fare for 1979 and (b) a 10 percent quarterly decrease in the average coach fare for 1979.

Figure 8 is a plot of the forecast results for 1979 by using the two alternative pricing scenarios. As we expect, a quarterly increase in fares results in a lower predicted demand than a quarterly decrease in fares, with the differences becoming more pronounced as the year goes on. However, due to the small magnitude of the fare coefficient, it is doubtful if this difference in forecasts is large enough to indicate significant impact of the alternative pricing strategies.

In summary, these results do not indicate a strong influence of fare alone on the demand for travel from mainland North America to Hawaii. Possible reasons for this include the fact that many holiday commitments are made many months in advance. In this case, fare may have a greater effect on demand after several periods. Furthermore, if fares to other vacation spots are increased concurrently, the incentive to shift to other destinations is reduced. Also, a large percentage of visitor travel to Hawaii is through organized tour agencies. The entire cost of the trip tends to be included in a single package, including air fare, hotel costs, and even entertainment expenses. Consequently, it is likely that demand for this type of travel is a function of several prices together rather than any single price component. Finally, it is possible that the demand for travel in this market is rather price inelastic. Many of these trips are once-in-a-lifetime experiences, and it is likely that other factors such as income or stage in the family life-cycle are operative. Consequently, these could outweigh the effect of air fare changes in determining the number of visitors who will travel to Hawaii.

SUMMARY AND CONCLUSIONS

To summarize our results, a univariate model of visitor travel from mainland North America to Hawaii is identified and estimated for the period from 1971 through 1978. The Box-Jenkins time-series analysis procedure is used. The resulting model is tested statistically against several alternative models and found to be preferable. It is further used to forecast travel from the mainland to Hawaii for 1979 and
1980. Although 17 of the 23 forecasts are not significantly different than the actual figures at the 95 percent level of confidence, the model tends to overpredict in general. This problem is alleviated substantially when the forecasts are updated each month with additional data points for 1979 and 1980.

In an attempt to add descriptive variables to the analysis, a bivariate time-series model is developed by using the average coach fare from mainland North America to Hawaii as the explanatory variable. The magnitude and significance of the fare parameter are low and the demand appears to be price inelastic. Alternative fare-pricing scenarios are studied, and their effect on forecasted demand is evident but not pronounced.

These results indicate that the Box-Jenkins methodology can be a useful tool in the analysis of an extensive time series of intercity travel demand. In cases where explanatory variables are poorly understood or where these data are unavailable, univariate analysis can result in a model that will produce useful short-term forecasts. Where a structural analysis is desired, explanatory variables can be added to the autoregressive components and transfer function models can be estimated. These are particularly useful to management and policy analysts who have some control over these variables. They can develop alternative future scenarios and study the effect these will have on future demand. Various elasticities of these variables with respect to demand can also be derived.

In terms of the Hawaii travel market, the bivariate model is a measure of the effect that fare alone has on demand. Research that uses a total visitor cost index would be useful in determining an overall cost elasticity of demand. Air carriers and other visitor industries could then determine their impact on this overall cost elasticity. Additional research might well be directed at the joint effect of price, economic activity, and changes in attractiveness of the destination market.

ACKNOWLEDGMENT

This paper is a result of research performed at Northwestern University through its Transportation Center and Department of Civil Engineering. The paper is an abridged version of Oberhausen's master's thesis, completed in February 1981. Koppelman served as Oberhausen's advisor throughout the course of the research.

Several organizations in Hawaii were responsible for providing the data necessary for the analysis, including the HVB, the Tax Foundation of Hawaii, and the Airports Division of the Hawaii Department of Transportation.

REFERENCES


Publication of this paper sponsored by Committee on Aviation Demand Forecasting.

Economic Justification of Air Service to Small Communities

JOHN HULET AND GORDON P. FISHER

This study is concerned with the allocation of air service to small communities (less than 50,000 population) at a time when the supply of that service is seriously diminishing and changing in character, especially since the Airline Deregulation Act of 1978. A quantitative methodology is developed as a tool for planning for short-haul service and establishes the minimum ridership required to justify the provision of air service. The model underlying the criterion takes into account two main factors: (a) the spatial separation of the community from a major hub and (b) the level of service offered at the nearest alternate airport and, if implemented, the local airport. The criterion equates the monetized time savings of local air service and the incremental costs to implement the service. This paper emphasizes the description of the trade-off mechanism between time and money by using classical cost elements of economic theory. A graphic analysis illustrates the validity of the functional shape of the disutility concept. The ultimate product of the methodology is an optimal configuration of local air service in terms of (a) link to be served, (b) airport investment level, (c) type of flight equipment, and (d) frequency of service.