Ross Walendziak – Final for Predictive Analytics Forecasting – Professor Rob Bradley—Fall 2021

**Forecasting Airport Footprints**

**Problem**

Accurately forecasting commercial air passenger volume is a difficult undertaking in that there are many dimensions to consider. At any given point in time air passenger volume could be affected by several factors including the long-term trend in air passenger volume, annual seasonal factors affecting the demand for air travel, changes in the macroeconomic landscape affecting cyclical demand for air travel as well as exogenous shocks like government enforced shutdowns of airports. Further complicating matters, these factors are often at work synchronously requiring simultaneous consideration of the net effect for a single time-period estimate. This paper will attempt to find a reasonable forecasting model to predict total monthly air passenger volume at Logan Airport in Boston Massachusetts.

**Significance**

This topic is worthy of study in that an accurate forecast of passenger volume at Logan Airport could significantly improve airline and airport staffing, city infrastructure planning around the airport, estimates of aggregate demand in the travel and leisure industry as well as serve as a good proxy for jet fuel demand and demand for other related petroleum products. Adequate staffing is important for airlines and airports to avoid travel delays and to maintain customer satisfaction levels. Although six-month forecasts of passenger volume are typically available by their nature, through advanced ticket sales data, longer term forecasts of passenger volume will allow airline and airport stakeholders to plan for the long term and invest in the appropriate training and technology programs to optimally process projected passenger volumes. By the same token, long-term forecasts of airport passenger volume will allow city and local government officials to upgrade ground transportation systems, moving patrons to the airport without delay. Analysts in the travel and leisure industry could find value in better understanding patterns in the trend and seasonal components of passenger volume. Is the long-term trend recently increasing or decreasing? Is the seasonal magnitude of air travel increasing or decreasing with time? Lastly, due to the large consumption share of highly refined petroleum products in the airline industry, forecasts for passenger volume could also serve as a proxy for jet fuel demand—having further forecasting potential for other highly refined products in the category such as kerosene or performance-oriented gasoline products.

**Data**

Historical data for Logan Airport passenger volume was sourced from Massport, the authority the operates Logan Airport. Specifically, monthly observation data from January 1999 through October 2021 was taken from the airport statistics section of the Massport website, located [here](https://www.massport.com/logan-airport/about-logan/airport-statistics/). On this webpage Massport publishes common airport statistics in individual monthly PDF documents. To extract total monthly passenger volume, this author manually saved all 274 documents to a local directory and subsequently built a function to load each document as a set of characters and scrape the relevant year and month and the associated total monthly passenger volume from each document. From there, a data frame was constructed consisting of two columns: month and passenger volume; and 273 rows: one for each unique month and year combination. To make the software cognizant of the time component, the data frame was converted to a tsibble object. Duplicates were detected for March 2002. Upon further investigation, the Massport website has March 2002 data duplicated and attached to the April 2002 hyperlink, thus the dataset is effectively missing an observation for April 2002. To proxy an observation representing April 2002, a linear interpolation was made between passenger volume observations in March 2002 and May 2002. Fortunately, no directional change in the seasonal component of the March, April and May sequence of months was detected during other years examined. As such, in this author’s opinion, a linear interpolation of the nearest months to April 2002 is a good estimate for the missing observation.

The resulting descriptive summary statistics are as follows:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Logan Airport: Descriptive Statistics for Total Monthly Passenger Volume** | | | | | | |
|  | Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max. |
| Monthly Passenger Volume (thousands) | 95.35 | 1,991.98 | 2,403.19 | 2,389.52 | 2,694.63 | 4,120.94 |

Chart, histogram

Description automatically generated

From the histogram, we can see the monthly observations are skewed left, placing the mean mildly below the median. We can also see emergence of large outliers at the extremes of the first quartile.

The raw time series is presented below with several noticeable basic features. First, there is a general upward growth trend throughout most of the data. Second, seasonality is a dominant factor through much of the data. Third, the time series shows some degree of cyclicality, deviating from the larger growth trend during the 2001 recession and the 2007 to 2008 Great Financial Crisis (GFC). Fourth, crisis is an important characteristic in the time series, with large drops in passenger volume during the September 2001 terror attacks and during the 2020 and 2021 Covid-19 (C19) global pandemic.

Chart

Description automatically generated

Season specific charts will allow to understand the seasonal and trend components more clearly. The left chart plots each year of passenger volume independently across months. From the plot was see that most years have an incremental increase over the prior year. We can also see that the year-over-year growth rate in passenger volume is greater in the in the 2013-2019 period than during the 1999-2012 period, represented by wider gaps between lines. This chart also shows that the magnitude of seasonality generally increases with time as the difference between peak and trough monthly passenger volume observations is larger for more recent years than it is for years further in the past. The exceptions to these trends are the years 2001, 2020 and 2021 – all representing crisis and instability. The chart on the right shows the time series specific to each month plotted through time, in black. The blue horizonal lines represent the average passenger volumes for each month across all observations. February has averaged the least amount passenger volume, with subsequent months showing incremental nonlinear gains in passenger volume through August—which seasonally represents the peak average monthly passenger volume. The months of September and October on average show a local trough and a local peak in the data set, respectively. November, December, and January show incremental declines in passenger volume from the local peak observed in September, ultimately converging with the global trough observed in February.

Chart

Description automatically generated Chart, timeline

Description automatically generated

Plots of autocorrelation and lags of monthly passenger volume quantitively confirm the seasonal and trend patters observed earlier through visual inspection. The left chart shows a strong linear relationship between the Kth observation of monthly passenger volume and the K-1 (1st lag) of the data. More intuitively, the February 2005 observation is highly predictive of the March 2005 observation. This first lag accounts for the strong upward growth trend exhibited through much of the data. The left chart also shows a strong linear relationship between the Kth and Kth-12 (12th lag) observations of monthly passenger volume. For example, the December 2013 observation is highly predictive of the December 2014 observation. This accounts for the annual seasonal pattern we observe in the data. The lag chart on right confirms the statistical significance of the linear relationships observed between the 1st and 12th lags in the data, with statistically significant relationships plotted above the dotted-blue line representing the 95% confidence interval.

Chart, scatter chart

Description automatically generated Chart

Description automatically generated

Lastly, we show a season and tread decomposition of the data with loss (STL) paired with a Box-Cox transformation using a lambda value of 0.12 (generally accepted as best for most monthly time series). The Box-Cox transformation can be understood as log transformation (month-over-month percentage changes) of the data when a lambda value of zero is used and a parallel shift in the data when a lambda value of one is used. This transformation, using a lambda value of 0.12, can intuitively be understood as closer to a log transformation, with some shape change. The result of the transformation is that our seasonal pattern becomes stationary, showing consistent seasonal magnitude through time—and enabling subsequent models to produce more confident forecasts of the data. Consistent seasonality was not observed in the raw data. This tread component is also easily understood in the chart below, showing consistent near linear growth since 2010. The remainder is centered around zero except for the September 2001 terror attack and the 2020 and 2021 C19 observations discussed earlier. These observations cannot be captured in either the trend or seasonal components of the data, thus their presence is in the remainder component.

Chart

Description automatically generated

**Literature**

In a 2001 paper in the Journal of Business & Economic statistics, Gregory, Smith and Yetman find that revisions to macroeconomic forecasts converge toward to the mean of the original set of forecasts. These finding are prevalent especially in forecasts of real output variables rather than economic measurements related to inflation or unemployment.[[1]](#footnote-1) They also find the mean forecast useful in establishing a consensus forecast when none of the other forecasts are alike (no mode forecast). As applied to this paper, airport passenger volume is a real output variable and using the mean of historical data points to make forecasts can serve as a consensus or base case forecast to assess other models.

A 1999 paper in the Journal of Economic Perspectives by Lloyd Thomas Jr found that forecasts of inflation using the Fischer model (a rational expectations model) underperformed a naïve model which in turn was outperformed by consumer survey data. The Fischer model forecasts inflation by taking the existing nominal interest rate and subtracting from it the expected real interest rate (assumed constant at 2%). The naïve model forecasts inflation using the most recent observation of the 12-month interest rate. The consumer survey data forecasts inflation using a collection of consumer responses on inflation expectations operated by the University of Michigan.[[2]](#footnote-2) The accuracy of models was assessed by calculating the RMSE of a 1-year ahead forecast of CPI from 1960 to 1997. This paper can use these findings by placing the final model selection into context. Perhaps a naïve model can outperform and the most rational expectation of continuation of the seasonal pattern of passenger volume. Additionally, perhaps the best fit model should be compared against simple survey done by the travel industry. Simply asking people how much they expect to travel by air in the Boston area could be superior to even the most sophisticated modeling techniques.

In a paper titled calendar effects of monthly time series: modeling and adjustment, Cleveland and Delvin show that monthly times series contain both a long-term trend and an annual seasonal component. The annual seasonal component can be further decomposed into a cyclical component representing seasonal patterns related to a particular month of the year and a calendar component related to the number of days in a month or the number of business days in a month.[[3]](#footnote-3) As it relates to this paper, passenger volume in any one month could be driven by cyclical seasonal demand for flight travel as well as simply the number of days in a month. Business travel on weekdays could be underrepresented in a month that has more days of weekend representation relative to other months. However, given that the seasonal components in this paper capture both the monthly cyclical nature of flight demand and calendar specific effects, it probably not necessary to separately consider each sub-component of seasonality.

In modeling and forecasting tourist flows to Barbados using structural time series models, Jackman and Greenbridge build a seasonal naïve model to serve as a benchmark for predicting tourist arrival volumes to Barbados using forecast horizons of one, four and eight quarters. In their estimation, the seasonal naïve model is necessary to assess the value of their more complex structural time series models which utilize the seasonal patterns in the tourist arrival data paired with multivariate use of value benchmarks—ranking relative prices in Barbados against comparable vacation alternatives is the US, Canada, and other Caribbean islands. The results across all forecast horizons show increased predictive accuracy, as assessed by the mean absolute percentage error (MAPE), of including the value component in additional to the seasonal component of the time series alone—as assess by the seasonal naïve forecast.[[4]](#footnote-4) The same type of seasonal benchmarking using a seasonal naïve model is applicable to this paper.

Pierce and Newbold, in the Journal of the American statistical association, show in a 1987 paper that seasonal time series can be fit using an ARIMA model that learns from separated trend and seasonal components of the series. The model then combines these forecasted component pieces to produce a robust model fit.[[5]](#footnote-5) In fact, to illustrate their findings, the authors apply their trend and cycle modeling to a series of logged monthly passenger totals at international airports, clearly topical to this discussion. This paper does not utilize ARIMA models, however, the decomposition model discussed later in this paper functionalizes some of the same ideas.

**Models**

To establish a set of base cases, four rudimentary models were built to serve as points of comparison for subsequent more advanced models. These basic models include a mean model, a linear trend model, a naïve model, and a seasonal naïve model. Conceptually, all four of these basic models are easy to grasp. The mean model calculates the arithmetic mean of monthly passenger volume across all observations and projects this value forward for any given periodic forecast. From our descriptive statistics, we know this would result in a monthly passenger estimate of 2,389,520 monthly passengers for any period forecast past September 2021. The linear trend model finds the best straight line fit across all observations, achieving such a straight line fit by minimizing the sum of squared residual errors. More commonly, the linear trend model is referred to as univariate simple linear regression. The naïve model is a simple as it gets, projecting forward the last observed value to estimate all forecast periods. The seasonal naïve method estimates the annual seasonal pattern by minimizing differences between the estimated and the actual observed values of the annual seasonal component of the time series. This method naively projects forward the most recent annual seasonal pattern to future forecast periods. In other words, the seasonal component of Jan 2022 is estimated to equal Jan 2021, Feb 2022 is estimated to equal Feb 2021, and so on.

For each model, assume that Box-Cox transformed data is fed into each model (as specified in the Data section of the report) and that forecasted values are comparable to the raw data set through application of the inverse Box-Cox function. To produce 2-year forecasts, the data was separated into a training set, consisting of monthly observations from Jan 1999 to Oct 2019, and a testing set, consisting of monthly observations from Nov 2019 through Oct 2021. These forecasts are presented in chart below. Clearly all the models have difficulty predicting the C19 event present in the 2-year test set. There is nothing in the historical training data set that would predict such a dramatic drop in passenger volume beginning in March 2020. However, C19 does serve as a check on the robustness of the model in fitting well to unforeseen circumstances. In the case of this forecast, we can see that the mean model is the best fit of the test data. However, the mean model seems to achieve these results simply by chance—biased low by very early observations in the time series. If an unexpected shock event in the test set saw a dramatic increase in passenger volume, then the mean model would be the worst performing model.

Chart

Description automatically generated

In effect, what we observe is a tradeoff between models that fit the training set well but do a poor job in predicting the C19 event in the test set – and models that fit the training set poorly but do an adequate job in fitting to C19. Below we see seasonal naïve and naïve models fit the training set well, nearly tracing the raw data (in black). In contrast, the mean model (in green) is a bad fit of the early and later observations of the training set. It also entirely ignores all seasonal characteristics of the data.

Chart

Description automatically generated

Ignorance of seasonal dynamics in the mean model is evident in the residual plots. Below we can see residuals generate a classic seasonal pattern. The trend is also ignored, as we see statistically significant autocorrelations of sequential lags in the acf plot.

Chart, histogram

Description automatically generated

The seasonal naïve model seems more reasonable for this data set, despite a poor fit to the test data in the forecast chart. However, the model is not perfect—residual errors from the model are plotted below. In the first comportment of the chart, we can see that the errors are correlated through time, consistently producing overestimates of the data during certain sub-periods, and consistently producing underestimates of the data during other sub-periods. This is because the trend is not factored in. We see this same idea captured in the statistically significant correlation between the kth residual and the K-1 through the K-9 residual lags. Notice however that the 12th lag of residual data is statistically insignificant, further confirming that the seasonal naïve model adequately captures the seasonality component. Further concerning is the distribution of error terms around zero not closing matching the normal distribution.

Chart, histogram

Description automatically generated

In the next section, we fit five more complex models to the data including: a decomposition model, a harmonic model with six Fourier terms, an exponential smoothing model with error, trend, and seasonal components, an Arima model and finally a piecewise linear model with Arima errors.

An improvement to the seasonal naïve model is the decomposition model. This model takes the seasonal naïve model in its current state and adds to it the trend component using seasonal and trend decomposition with losses (STL). To account for the increase in the trend over the past 10 years, the trend component is estimated using only the most recent 60-month (5-year) window in the training data.

Residuals from the decomposition model are plotted below. Here we see that residuals are much more evenly distributed around zero and are uncorrelated through time. In the decomposition model, much of the information embedded in the statistically significant near-term residual lags from seasonal naïve model has been extracted and put towards a better model fit by incorporating the trend component. We also see that the residuals from the decomposition model produce a more normal distribution centered around zero as compared to the seasonally naïve model.

Chart, box and whisker chart

Description automatically generated with medium confidence

Our harmonic regression model uses the trend over the full training data set plus six Fourier terms to capture the seasonality in the data. Fourier terms can be best understood as a series of Sin and Cos functions with coefficients placed in front of the terms to model the seasonal cycles present in the data. Broadly speaking, more Fourier terms gives the model more flexibility and degrees of freedom to achieve a good fit to the data. The ETS model represents an exponential smoothing technique where the best fit to our data uses an additive error term, an additive seasonal term, and no trend component. The ETS model fit equation is Model(ETS(passengers ~ error(“A”) + trend(“N”) + season(“A”))). That is, the ETS model is specifically an ETS(A,N,A) model. The Arima model that best first the passengers data is an ARIMA(0,1,1)(0,1,1)[12] model. The model is understood in the framework ARIMA(p,d,q)(P,D,Q) terminology. The “p” component represents the number of autoregressive lags used to best fit the model. The “d” component represents the degree of differencing using to establish stationary data for modeling, where differencing is understood as . The “q” term represents the order of the moving average term (lags of the error term) to fit the model. Nonseasonal versus seasonal autoregressive lags, differencing, and moving average components are differentiated with the use of lowercase and capitalized “p,d,q” terms, respectively. For example, one monthly seasonal difference of the data, represented by D=1, could be understood as . The suffix ([12]) appended to the end of the model represents modeling for a monthly timeseries. Plots of the residuals from the ETS and ARIMA models are shown below. The two are remarkably similar, showing distributions centered around zero and statistically insignificant ACF lags.

Chart, box and whisker chart

Description automatically generatedA picture containing chart

Description automatically generated

The last model to consider is a piecewise linear fit using Arima errors. Here, the model can be thought of in two distinct parts. First, a linear piecewise regression is fit using nodes at crisis points in the training data set. In this case, a piecewise fit is done using nodes Sept 2001 and Jan 2002—representing the beginning and end periods of the 2001 September 11th terror attacks. The piecewise linear model is fit, and residuals are produced. Subsequently, an Arima model is fit to error terms, thus extracting further predictive power from errors present in the piecewise linear fit. In this case an ARIMA(3,0,0)(1,0,0)[12] model best fits the piecewise residuals.

The graph below shows the fitted values of each more advanced models previously discussed plotted against the actual observed values in the training set (in black). Overall, we notice all models include both seasonality and some representation of a trend component. As such, all fitted values are similar. Visually, we can see that the harmonic K6 model produces consistent overestimates and then underestimates of the time series. Apart from that, it is difficult to distinguish one model from the next. All seem like plausible candidates for best fit. We will leave quantitative evaluation of the models to the performance section of the report.

A picture containing chart

Description automatically generated

Given the similarity of fit to training data, it makes sense that the 2-year forecasted values of these model also produce similar results. Below is 2-year forecast plotted against the test set in (in black). Like the basic models, the worst fitting advanced model to historical data (harmonic\_K6) produces the best overall estimate of the downturn in passenger volume from C19. Generally, the models predict very similar forecast values.

Chart, line chart

Description automatically generated

**Performance**

The root mean squared error (RMSE) was used to evaluate the performance of all nine models discussed under three different methodologies. First, the training RMSE was calculate based on how well fit each model was to the training data set. Second, the testing RMSE was calculated by exposing the models to the training data set and evaluating the performance based on errors produced in forecasting the 2-year test data set. Third, the RMSE was calculated using a cross validated (CV RMSE) approach. Starting with an initial bank of 36 training observations (3-years) the RMSE was calculated on a 2-year forecast. Each subsequent iteration of the CV approach added 3 months of observations to the initial 36 observation training bank and again produced a 2-year forecast beginning at the end of the new training bank. This process completes when reaching the end of the full time series. The RMSE of each iteration was averaged, producing the CV RMSE for each model in the table above. The most accurate model under each scenario (train, test and CV) is highlighted in green. The second and third most accurate model under each scenario are highlighted in blue and orange, respectively.

Traditionally, the CV RMSE is used for best model selection, as it represents the most robust approach for assessing models under test conditions across full spectrum of regime environments over time. However, given the unique circumstances that C19 presents—I think this traditional method of assessment overly penalizes models that underperform during the C19 period. In the very long term, we should not expect C19 to be a repeatable event. Given that the goal of this paper is to produce a 2-year forecast of passenger volume, this time horizon represents a middle ground, where the affects of C19 will inevitably affect passenger volume in the nearest months—but C19 realistically should be expected to have muted effects on passenger volume 12, 18 and 24 months in the future.



As such, this author uses a relative model ranking system based on RMSE under the three scenarios for best model selection. Under each evaluation scenario, the models where ranked (1 through 9). A simple average of the rankings was taken for each model, establishing a more wholistic assessment. Based on this criteria, the best performing model is the ETS(A,N,A). In this model we can see it fits the training data well, scoring only a few points higher than the ARIMA model, which scored best in this condition. It performs in the middle of the pack in forecasting the C19 event present in the test set. Importantly, it scores second best in the cross validated approach. Intuitively, we can understand the ETS(A,N,A) model as fitting the historical data slightly less well that the optimal ARIMA model, but being more robust to unexpected negative shocks to passenger volume that new variants of Covid-19 may present.

**Final 2-year Forecast**

Finally, the ETS(A,N,A) model is trained on all available data, representing the full set of observations from Jan 1999 to Oct 2021. A two-year forecast for the Nov 2021 through Oct 2023 period is presented below. The 95% confidence interval is presented in the lighter shade of purple. The 80% forecast estimate is presented in a darker shade of purple. The mean forecast is shown in the blue point estimate line plot. Here we see more uncertainty in the upside growth and recovery of passenger volume from C19 lows than in the potential downside negative shock to future passenger volume. Given that the future use of airlines by the public remains highly uncertain, and travel alternatives such as video conferencing are in their infancy—the 2-year forecast below presents a realistic path forward—maintaining the seasonal patterns observed historically, but at a more modest levels than what was experienced during peak travel periods throughout the 2017, 2018 and 2019 years.

Chart, line chart

Description automatically generated

1. Gregory, Allan W., et al. “Testing for Forecast Consensus.” Journal of Business & Economic Statistics, vol. 19, no. 1, [American Statistical Association, Taylor & Francis, Ltd.], 2001, pp. 34–43, http://www.jstor.org/stable/1392539. [↑](#footnote-ref-1)
2. Thomas, Lloyd B. “Survey Measures of Expected U.S. Inflation.” The Journal of Economic Perspectives, vol. 13, no. 4, American Economic Association, 1999, pp. 125–44, http://www.jstor.org/stable/2647016. [↑](#footnote-ref-2)
3. Cleveland, William S., and Susan J. Devlin. “Calendar Effects in Monthly Time Series: Modeling and Adjustment.” Journal of the American Statistical Association, vol. 77, no. 379, [American Statistical Association, Taylor & Francis, Ltd.], 1982, pp. 520–28, https://doi.org/10.2307/2287705. [↑](#footnote-ref-3)
4. Jackman, Mahalia, and Kevin Greenidge. “Modelling and Forecasting Tourist Flows to Barbados Using Structural Time Series Models.” Tourism and Hospitality Research, vol. 10, no. 1, Sage Publications, Ltd., 2010, pp. 1–13, http://www.jstor.org/stable/23745508. [↑](#footnote-ref-4)
5. George E. P. Box, et al. “Estimating Trend and Growth Rates in Seasonal Time Series.” Journal of the American Statistical Association, vol. 82, no. 397, [American Statistical Association, Taylor & Francis, Ltd.], 1987, pp. 276–82, https://doi.org/10.2307/2289164. [↑](#footnote-ref-5)