FORECASTING AIRLINE PASSENGER MILES

Lara O'Brien, United Airlines, Department of Finance,
Houston, TX, larakobrien@gmail.com

C. Barry Pfitzner, Department of Economics and Business,
Randolph-Macon College, Ashland, VA, bpfitzne@rmc.edu

ABSTRACT

This project utilizes classical forecasting techniques to predict international revenue passenger miles for U.S. air carriers. The data are monthly and seasonally unadjusted. The approach we employ involves stabilization of the variance of the series, trend fitting, seasonal adjustment through the use of the dummy variable method, and the autoregressive moving average (ARMA) cyclical representation. Forecasts are presented for the period following the end of the dataset. These forecasts mimic the time series properties of the air miles series very well. We produce forecasts in real time for the last five years of the data set and find that the model performs well as judged by traditional forecast measures.

INTRODUCTION

International air passenger travel was interrupted by the events of September 11, 2001. It is somewhat surprising to learn how quickly this series of international revenue passenger miles get back to trend following that interruption. Forecasting such a series is clearly important to air carriers as they attempt to predict utilization, costs, input requirements and the like for future years and months, but these methods may also be useful for other policymakers as they attempt to predict the requirement for other services and to assess the impact of other causes of disruptions in service.

DATA

For this project monthly measures of international air revenue passenger miles of US carriers were collected for the range, January 1996 to November 2011. The source of the data is the Research and Innovative Technology Administration of the Bureau of Transportation Statistics [5].

The series is shown in Figure 1. Several time series properties are easily observed from the figure. First, the series displays a mild (though significant) upward trend. Second, the repetitive seasonal nature of the data is obvious. Third, the effects on the series of the events of September 11, 2001 are clear with the sharp drop in the series at that date. Finally, we also observe that the variance of the series is positively related to its level.

METHOD AND ESTIMATION RESULTS

Transformation

The last observation from the previous paragraph led us to consider modeling in either the square root or the natural log of the series. These are the two most common methods for transforming a time series so that the variance is relatively constant. Figure 2 is the series in (natural) log form. The time series properties described above are still evident on Figure 2 with the exception that the variance of the series is now relatively constant. The square root transformation (not depicted) did not stabilize the variance of the series as well as the log transformation.

Figure 1: International Revenue Passenger Miles for the U.S.

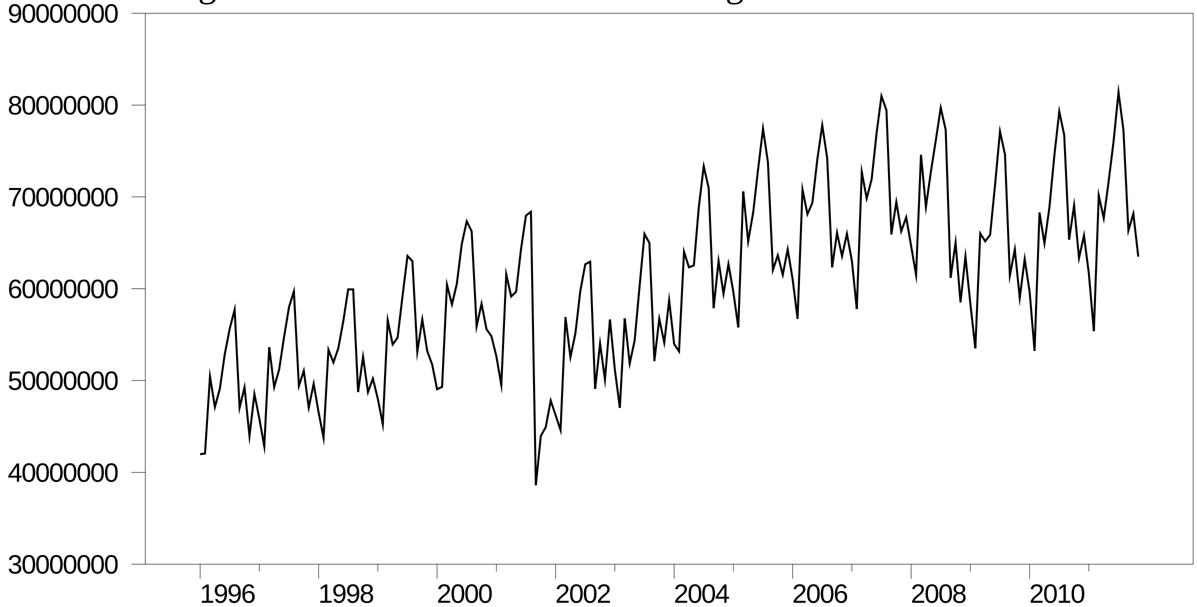
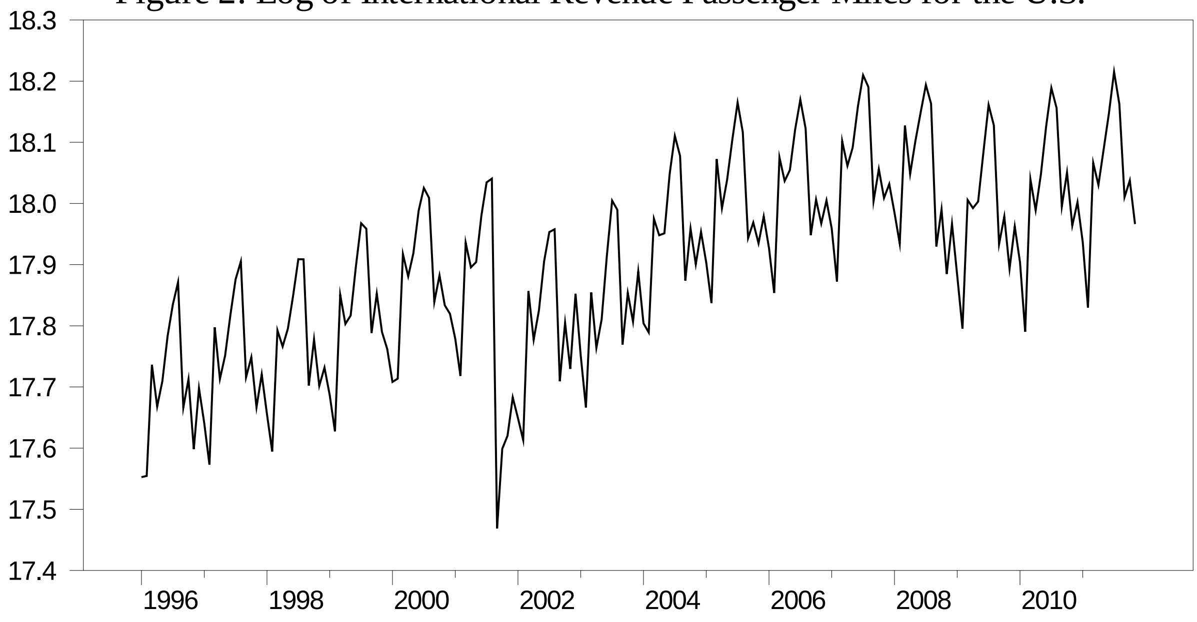


Figure 2: Log of International Revenue Passenger Miles for the U.S.



Choosing a Trend

The second step in modeling the series is to choose a trend to fit to the data in log form. We entertained a simple linear trend (though in logs, it's often called a log-trend) and a quadratic trend in logs as well.

Standard complexity penalized model selection criteria were employed in choosing the remainder of the modeling. The two standard penalized likelihood selection criteria are the Akaike information criterion *(AIC)* and the Schwarz information criterion *(SIC)* represented as follows:

$$AIC = (2k/T) + log(\sigma^2) \tag{1}$$

$$SIC = [k\,log(T)/T] + log(\sigma^2), \tag{2}$$

where k is the total number of estimated coefficients in the equation, T is the number of usable observations, and σ^2 is the scalar estimate of the variance of the equation's disturbance term. In this particular case the *AIC* and the *SIC* are very nearly identical for the simple trend and the quadratic trend. We choose the more parsimonious simple trend for further modeling.

Also included in the trend model (and all subsequent estimations) is a "pulse" dummy variable to account for the drop in the series that occurs in September of 2001. We choose a pulse dummy because the series returns to trend relatively quickly and (as indicated later) the forecasting equation will eventually include autoregressive and moving average terms. It could, of course, be argued in favor of other dummy variable representations for the 9-11 effect.

Modeling the Seasonality

The seasonality of the series is modeled via a set of dummy variables. Eleven dummy variables (one fewer than the number of months, since each equation includes an intercept) were created and included in the regression.

Though the data are clearly seasonal by casual observation, we nonetheless also relied in the standard F-test and the values of the *AIC* and *SIC* to determine whether or not the series was subject to seasonal variation. As anticipated, a null hypothesis of non-seasonality was rejected resoundingly for the air miles data. (These results are available from the authors on request.)

Figure 3 depicts a fit to the data including the simple trend, the seasonal dummies and the 9/11 dummy variable.

Modeling Cycles

Notice that in Figure 3, the trend and seasonality are clearly well modeled, but the fitted values do not "cling" to the series when the series moves above or below the fitted values. That indicates that there are cycles in the data. A visual examination of the residuals from the estimation with trend and seasonality variables entered in the regression also reveals that the residuals exhibit autocorrelation—confirming the presence of cycles.

Examination of the autocorrelations and partial autocorrelations of the residual series reveals that the autocorrelations "tail off" and the partials "cut off" after lag 1. This behavior is indicative of a first order autoregressive representation, often called an AR(1), of the cyclical nature of the series. Thus we choose to add to the model that additional parameter estimate. Needless to say, the first-order autoregressive term is highly significant statistically.

The software we used in this research also allow for "automatic" choice for modeling cycles in the residual series. Using that procedure, based on the *AIC* and *SIC*, a model with the first-order autoregressive *and* a first-order moving average term are chosen. We also use that model for forecasting with somewhat improved results. The graph of that model is nearly identical to figure 4, and thus not displayed.

Figure 3: Simple Trend with Seasonals

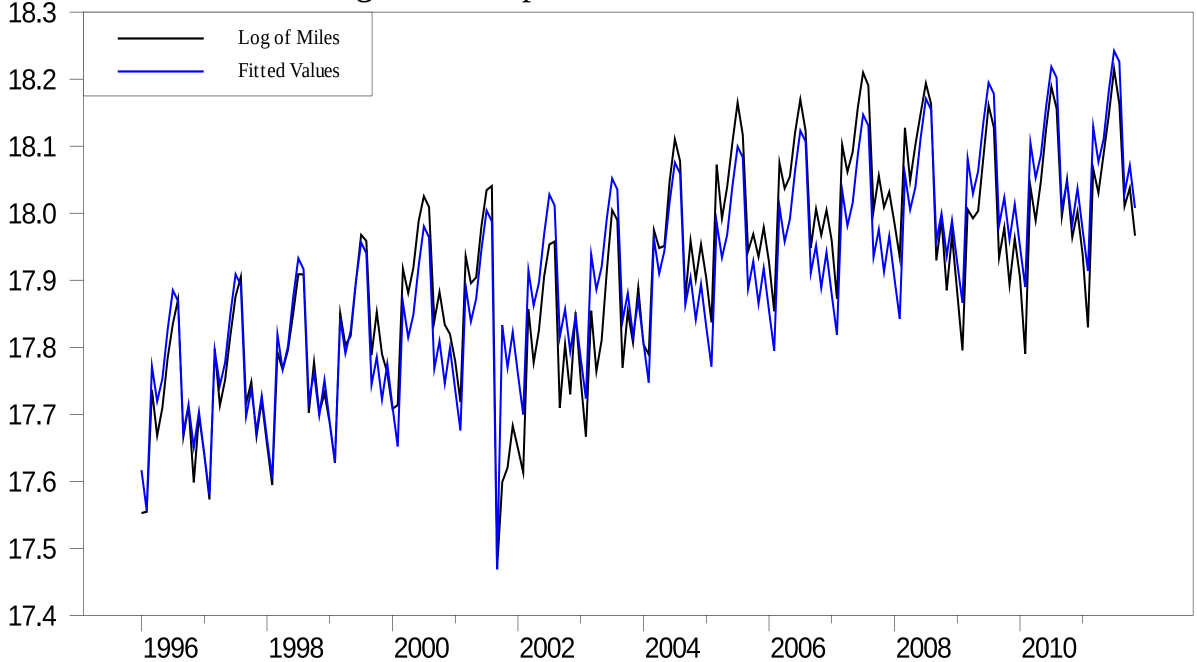
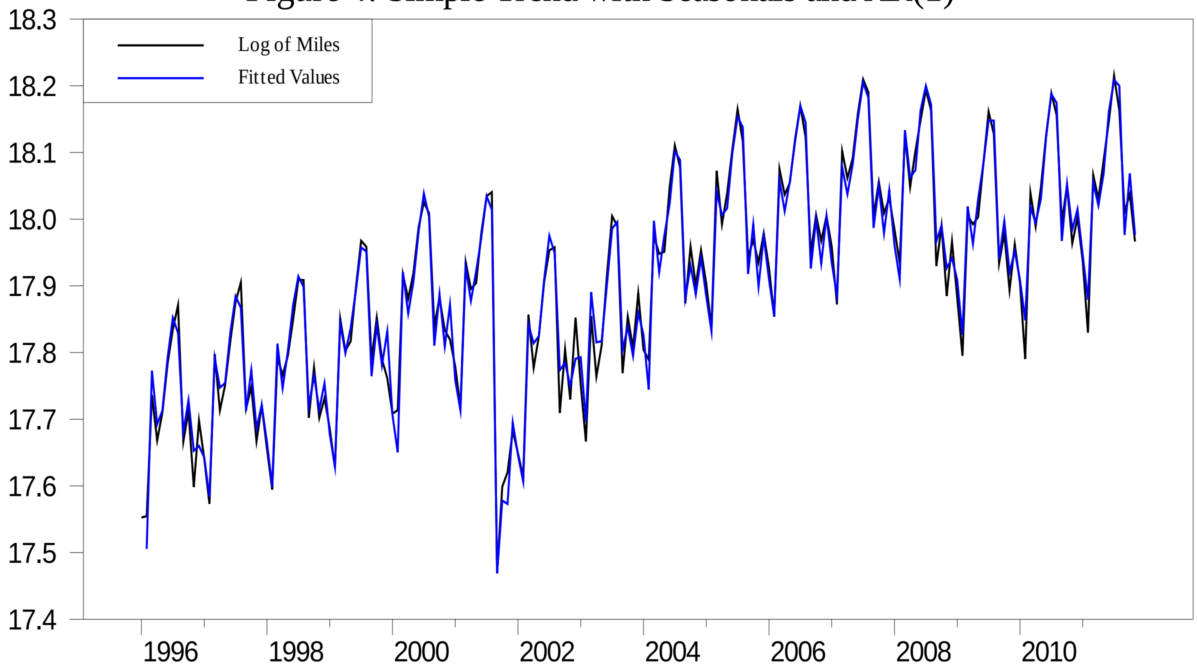
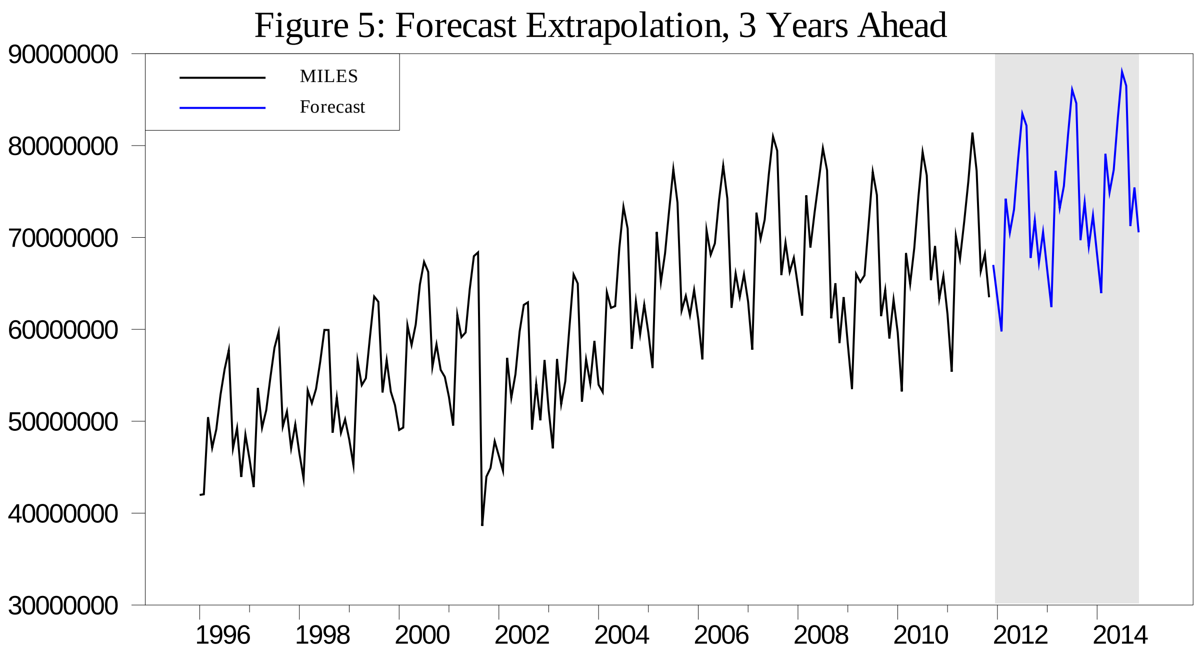


Figure 4: Simple Trend with Seasonals and AR(1)



FORECASTING RESULTS

As a first exercise in forecasting, we produced forecasts extending past the known data. These forecasts cannot be judged for accuracy (until time passes), but they should exhibit the times series properties of this data set. Figure 5 is the graph of the actual data series with the forecasts represented in the shaded area past the end of the data set. Note here that the transformation into natural logs has been "undone" by taking the anti-logs of the series and the forecasts.



In order to judge the forecast accuracy of a model, it is usually advisable to generate forecasts in "real time" for known values of the series. The idea is to estimate the model using a sub-sample of the data less than the full sample, produce forecasts from that model and compare the forecasts to the actual values. Here, we estimate the model from the beginning of the data through December of 2006 and then produce forecasts for twelve steps ahead (January 2007 through December 2007). Then the model is re-estimated, adding January 2007 to the estimation period, forecasts are generated for February 2007 through January 2008. This method is called sequential updating. In this way we are able to produce 60 one-step ahead forecasts, 59 two-step ahead forecasts, and so forth. These forecasts can then be compared to the actual values of the series to assess forecast accuracy.

The results of that exercise are summarized in Table I. Forecasts at each horizon are compared to the actual values by means of standard measures of forecast accuracy. The definitions of these statistics are as follows:

$$Mean\ error = \frac{\sum(actual - forecast)}{T} \tag{3}$$

$$Mean\ absolute\ error = \frac{\sum|actual - forecast|}{T} \tag{4}$$

$$Root\ mean\ squared\ error = \sqrt{\frac{\Sigma(Actual - forecast)^2}{T}} \tag{5}$$

$$Theil's\ U = \frac{RMSE\ model}{RMSE\ naive} \tag{6}$$

Where T is the number of forecast periods, and the RMSE naïve represents the forecasts of no change in the series.

The mean error at one step ahead is approximately zero (the value of the series is about 18 working in logs, so the mean error of -0.0022739 working in logs is only -0.0001263 as a proportion). Also, at one step ahead, the *RMSE* is about .02161, or a little over one-tenth of one percent of the mean value of the series. That is very small. Theil's U indicates that this forecasting model is considerably more accurate than a naïve forecast.

As expected the forecast errors are larger the farther ahead you forecast in general and Theil's U shows that the forecasts from the model are superior to the last known value of the series. Theil's U increases considerably at 12 steps ahead. This is because there is not much trend in this dataset, and there is a significant amount of seasonality. For this reason, the actual value twelve steps ahead is generally very close to value twelve months prior. However, a Theil's U of .90 still considerably outperforms the naïve forecast.

Table I: Forecast Error Statistics, AR Model

Steps Ahead	*Mean Error*	*Mean Absolute Error*	*RMS Error*	*Theil's U*	*Obs*
1	-0.0022739	0.0176696	0.0216105	0.2125	60
2	-0.0041244	0.0216662	0.0259653	0.2132	59
3	-0.0057847	0.0252462	0.0303724	0.2157	58
4	-0.0073922	0.0296147	0.0348217	0.2169	57
5	-0.0089098	0.0328857	0.0385220	0.2268	56
6	-0.0107532	0.0343536	0.0404595	0.2123	55
7	-0.0124804	0.0370219	0.0428528	0.2503	54
8	-0.0142906	0.0378767	0.0440188	0.2671	53
9	-0.0160589	0.0389867	0.0452091	0.3123	52
10	-0.0178693	0.0404481	0.0463359	0.3661	51
11	-0.0196736	0.0410867	0.0467377	0.4199	50
12	-0.0214184	0.0414010	0.0471722	0.9024	49

Recall that the "automatic fit" for cycles chose a cyclical model with a first order regressive term and a first order moving average term, that is, an ARMA(1,1). We estimated that model as well. It is interesting to note that while the moving average term does not meet strict tests of statistical significance, the estimation does produce better forecasting results. Table II presents the same forecast statistics for the model with a first-order moving average term included. For most forecast horizons, the forecast improvements for the ARMA model are small, but at 12 steps ahead (one year) the forecasts are almost 7% more accurate. We do not have a ready explanation of that particular result.

Table II: Forecast Error Statistics, ARMA Model

Steps Ahead	*Mean Error*	*Mean Absolute Error*	*RMS Error*	*Theil's U*	*Obs*
1	-0.00159	0.017257	0.021093	0.2074	60
2	-0.00270	0.019996	0.024586	0.2018	59
3	-0.00392	0.022945	0.027927	0.1984	58
4	-0.00518	0.025938	0.031590	0.1968	57
5	-0.00636	0.028840	0.034690	0.2043	56
6	-0.00788	0.029886	0.036413	0.1911	55
7	-0.00929	0.032957	0.039326	0.2297	54
8	-0.01085	0.034156	0.040668	0.2468	53
9	-0.01242	0.035177	0.041964	0.2899	52
10	-0.01404	0.036377	0.043374	0.3427	51
11	-0.01571	0.036859	0.043743	0.3930	50
12	-0.01753	0.037056	0.043987	0.8415	49

Forecasts can be tested for "optimality" with respect to the data from which the forecasts were generated. This test is known as a Mincer-Zarnowicz (M-Z) [5] regression:

$$y_{t+h} = \beta_0 + \beta_1 y_{t+h,t} + u_t \tag{7}$$

where y is the forecast variable, h is the number of steps ahead, t is the current time period, and u_t is the white noise error term. The test for optimality is then the joint hypothesis that $(\beta_0, \beta_1) = (0, 1)$. The test is whether or not the forecasts are equal, on average, to the realized values of the series. The results of the M-Z test for the ARMA model at the one-step ahead horizon indicate that the model is optimal in the sense of Mincer and Zarnowicz. The calculated value of $F_{(2, 58)} = 0.453$, with a p-value of 0.64, indicates that the estimates of β_0 and β_1 do not differ from 0 and 1 respectively.

CONCLUSIONS

This paper takes a classical approach to modeling and forecasting international airline passenger miles for US carriers. We identify trend, seasonal, cyclical components, and model the effect of the events of September 11, 2001 as a dummy (intervention) variable. The model performs well in terms of traditional forecast statistics, showing significant superiority to a naïve forecasting standard. The model also passes easily the Mincer-Zarnowicz test of optimality.

Forecasts such as those produced here can be adapted to, and be useful for, predicting individual carrier demand, airport use, personnel demand, and many other activities related to air travel. Models such as the one produced here are also useful in forecasting seasonal, cyclical, and long-term levels of variables of interest. Such estimations are also valuable in predicting the effects, both short-term and long-term, of interruptions in the series. In the case of international air travel, it is perhaps surprising how quickly the series seemed to have returned to trend following an unprecedented interruption.

REFERENCES

[1] Enders, Walter. *Applied econometric time series,* John Wiley & Sons, INC. New York, 2004.

[2] Diebold, Francis X., *Elements of forecasting*, 4th edition, Thomson, South-Western Publishing, Cincinnati, OH, 2007.

[3] Doan, Thomas. *RATS user's guide*. Estima, Evanston, Illinois, 2010.

[4] Mincer, J. and Zarnowicz. "The evaluation of economic forecasts," in J. Mincer (ed.) *Economic forecasts and expectations.* 1969.

[5] Source for the data: http://www.bts.gov/xml/air_traffic/src/datadisp.xml