Gregory Buisson

Forecast Analysis of Air Travel Before and After 9/11

Introduction and Business Understanding:

Passenger travel patterns are shaped by numerous external variables, some predictable like seasonal trends, and others unforeseen, such as the September 11, 2001, attacks. These unexpected events can significantly disrupt travel patterns, resulting in deviations from prior forecasts. This study focuses on evaluating various forecasting approaches to better understand and measure the impact of the 9/11 events on long-distance commercial airline travel behavior.

Data Understanding:

The pre-September 2001 time series was plotted, showing seasonality after adjusting lambda to 0. In total there are 172 observations, however, 32 of these observations are considered post-September 2001 and are excluded in some models. To account for seasonality, two columns were added, month number and season. Season is a nominal variable, creating 12 dummy variables for the season. The main variables of interest are air RPM, season, and month number which are then used to find predicted formula air RPM and ARIMA forecast.

Analysis:

Regression Modeling:

A regression model was then created using air rpm with two variables: the month number and season. Both variables are significant. A prediction formula was then saved from the new model and then implemented into the line graph seen in Appendix D along with the actual time series data and best ARIMA model.

ARIMA:

The ARIMA model does not use data from September 2001 to April 2004, so they were excluded. The model was created using time series analysis with month as the time and time series being air RPM, as seen in Appendix A. The autocorrelation function (ACF) shows lag-1 to lag-6 values tapering to a minimum while lag-6 to lag-12 values increase to a maximum. This is consistent with the 12-month seasonal pattern. The partial autocorrelation function (PACF) showed a large peak at lag-1 and lag-7 with smaller peaks around lag-12. The prominent peak at lag 1 in the PACF indicates that an AR(1) model might be suitable. However, given the strong seasonal and AR(1) effects, the researcher will take both seasonal and first differences to determine if the autocorrelation is resolved. The results of this approach, using the ARIMA(0,1,0)(0,1,0)12 model, are detailed in Appendix B. The ACF and PACF still exhibit peaks at lags 1 and 12, though these are less pronounced compared to the original data. This suggests the persistence of a seasonal and trend component, indicative of either an MA or an AR element. To rectify this the following models were fit: ARIMA(0,1,1)(0,1,1)12, ARIMA(1,1,0)(1,1,0)12, ARIMA(1,1,0)(0,1,1)12, ARIMA(0,1,1)(1,1,0)12, and ARIMA(1,1,1)(0,1,1)12. The best fit model was ARIMA(1,1,1)(0,1,1)12 with a R-squared value of 0.98 (Figure 1).

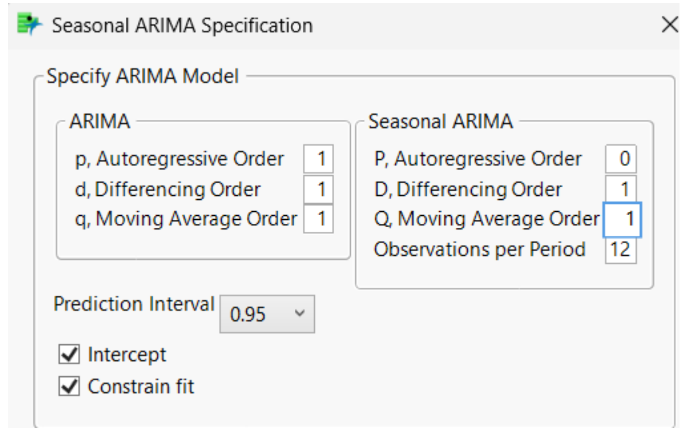


Figure 1: JMP Specifications for Best Model

The other models showed smaller R-squared values and some showed an insignificant seasonal component of AR. Appendix C presents the final statistics for the model. The ACF and PACF plots of the residuals reveal no significant values.

Actual versus Forecast:

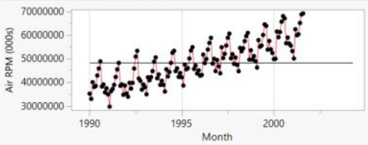
Forecast values were compared to actual values between September 2001 to April 2004 in Excel. Air RPM demonstrates a good fit beyond 2001. The pre-September 2001 average error was 0.3% while the post-September 11 forecast error was -4.4%. Much of this error occurred between October 2001 and May 2003.

Conclusion:

The model chosen for analyzing AIR RPM was an ARIMA(1,1,1)(0,1,1)12. This model featured three statistically significant parameters and exhibited a high adjusted R-squared value. Importantly, the residuals of this model showed minimal autocorrelation. The superior accuracy of the ARIMA model over ordinary regression estimation in time series analysis is primarily due to its ability to account for dependencies between observations. In time series data, the value of a current observation is often influenced by the values of previous observations, a relationship known as autocorrelation. ARIMA models are specially designed to capture this aspect by considering how past data points, and the errors associated with these points, influence future values. In contrast, ordinary regression models generally operate under the assumption that each data point is independent of the others. This assumption often doesn't hold in time series analysis, where past values significantly influence future ones. As a result, while traditional regression models may overlook the time-dependent structure inherent in the data, ARIMA models embrace and utilize this structure. By adjusting predictions based on how the data has been trending, ARIMA models provide more accurate forecasts for time series data, making them particularly valuable in situations where understanding and predicting temporal dynamics are crucial.

Appendix A: Time Series, ACF and PACF for Air RPM between January 1990 and August 2001

Time Series Air RPM (000s)



Mean	48213758
Std	8660671.3
N	140
Zero Mean ADF	0.0715757
Single Mean ADF	-3.088021
Trend ADF	-6.350581

Time Series Basic Diagnostics

Lag	AutoCorr	Ljung-Box Q	p-Value	Lag	Partial
0	1.0000	.	.	0	1.0000
1	0.8221	96.6640	<.0001*	1	0.8221
2	0.7142	170.143	<.0001*	2	0.1182
3	0.6133	224.722	<.0001*	3	-0.0051
4	0.5231	264.716	<.0001*	4	-0.0149
5	0.4549	295.191	<.0001*	5	0.0242
6	0.3251	310.869	<.0001*	6	-0.2134
7	0.4484	340.917	<.0001*	7	0.6798
8	0.5075	379.705	<.0001*	8	0.0278
9	0.5705	429.099	<.0001*	9	0.1440
10	0.6396	491.651	<.0001*	10	0.1737
11	0.6907	565.180	<.0001*	11	0.1799
12	0.8121	667.598	<.0001*	12	0.2363
13	0.6512	733.972	<.0001*	13	-0.5043
14	0.5477	781.297	<.0001*	14	-0.0614
15	0.4467	813.031	<.0001*	15	-0.1473
16	0.3550	833.240	<.0001*	16	-0.0395
17	0.2911	846.935	<.0001*	17	-0.1314
18	0.1749	851.917	<.0001*	18	0.0345
19	0.2887	865.606	<.0001*	19	-0.0554
20	0.3409	884.865	<.0001*	20	-0.0397
21	0.3946	910.877	<.0001*	21	0.0146
22	0.4549	945.739	<.0001*	22	0.0492
23	0.4948	987.333	<.0001*	23	0.1143
24	0.5977	1048.55	<.0001*	24	0.0762
25	0.4424	1082.38	<.0001*	25	-0.1752

Appendix B: Seasonal and First Differences

Model: Seasonal ARIMA(0, 1, 0)(0, 1, 0)12

Model Summary

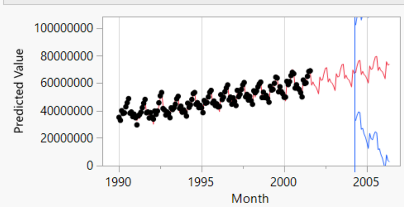
DF	126	Stable	Yes
Sum of Squared Innovations	2.8188e+14	Invertible	Yes
Sum of Squared Residuals	2.8188e+14		
Variance Estimate	2.2372e+12		
Standard Deviation	1495720.61		
Akaike's 'A' Information Criterion	3972.80857		
Schwarz's Bayesian Criterion	3975.65276		
RSquare	0.96883726		
RSquare Adj	0.96883726		
MAPE	2.48674772		
MAE	1151409.48		
-2LogLikelihood	3970.80857		

Parameter Estimates

Term	Factor	Lag	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	1	0	18774.737	129781.8	0.14	0.8852

Constant Estimate	Mu
18774.7374	18774.7374

Forecast



Residuals

Lag	AutoCorr	Ljung-Box Q	p-Value	Lag	Partial
0	1.0000	.	.	0	1.0000
1	-0.1623	3.4271	0.0641	1	-0.1623
2	-0.2129	9.3693	0.0092*	2	-0.2458
3	-0.0919	10.4850	0.0149*	3	-0.1911
4	-0.0407	10.7057	0.0301*	4	-0.1761
5	0.0360	10.8800	0.0538	5	-0.0993
6	0.0055	10.8841	0.0920	6	-0.0976
7	0.0267	10.9814	0.1394	7	-0.0420
8	0.0594	11.4669	0.1766	8	0.0323
9	-0.0129	11.4900	0.2436	9	0.0122
10	0.0874	12.5591	0.2494	10	0.1475
11	0.0391	12.7748	0.3083	11	0.1634
12	-0.4854	46.3344	<.0001*	12	-0.4291
13	0.1136	48.1903	<.0001*	13	-0.0669
14	0.1705	52.4033	<.0001*	14	-0.0355
15	0.0292	52.5282	<.0001*	15	-0.0937
16	0.0427	52.7976	<.0001*	16	0.0129
17	-0.1243	55.0972	<.0001*	17	-0.1173
18	-0.0496	55.4673	<.0001*	18	-0.1316
19	-0.0151	55.5019	<.0001*	19	-0.0934
20	0.0528	55.9287	<.0001*	20	0.0250
21	0.0005	55.9287	<.0001*	21	-0.0871
22	-0.0099	55.9441	<.0001*	22	0.0251
23	-0.0383	56.1748	0.0001*	23	-0.0328
24	0.1367	59.1491	<.0001*	24	-0.1657
25	-0.0270	59.2663	0.0001*	25	-0.0014

Appendix C: ARIMA(1,1,1)(0,1,1)12

Model: Seasonal ARIMA(1, 1, 1)(0, 1, 1)12

Model Summary

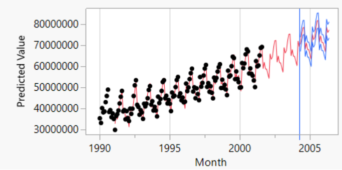
DF	123	Stable	Yes
Sum of Squared Innovations	1.6208e+14	Invertible	Yes
Sum of Squared Residuals	1.7828e+14		
Variance Estimate	1.3177e+12		
Standard Deviation	1147930.93		
Akaike's 'A' Information Criterion	3917.86511		
Schwarz's Bayesian Criterion	3929.24186		
RSquare	0.98029141		
RSquare Adj	0.97981071		
MAPE	1.93974457		
MAE	908789.613		
-2LogLikelihood	3909.86511		

Parameter Estimates

Term	Factor	Lag	Estimate	Std Error	t Ratio	Prob>|t|
AR1,1	1	1	0.621	0.072	8.64	<.0001*
MA1,1	1	1	1.000	0.045	21.98	<.0001*
MA2,12	2	12	0.561	0.075	7.47	<.0001*
Intercept	1	0	11432.973	4360.484	2.62	0.0098*

Constant Estimate	Mu
4334.5679	11432.9732

Forecast



Residuals

Lag	AutoCorr	Ljung-Box Q	p-Value	Lag	Partial
0	1.0000	.	.	0	1.0000
1	0.0154	0.0309	0.8605	1	0.0154
2	-0.0818	0.9087	0.6348	2	-0.0821
3	-0.0686	1.5300	0.6754	3	-0.0664
4	0.0569	1.9618	0.7428	4	0.0527
5	0.0640	2.5125	0.7746	5	0.0524
6	0.0154	2.5445	0.8635	6	0.0182
7	0.0461	2.8343	0.8999	7	0.0629
8	0.1070	4.4099	0.8184	8	0.1154
9	-0.0470	4.7166	0.8583	9	-0.0459
10	0.0009	4.7167	0.9093	10	0.0212
11	-0.0414	4.9582	0.9332	11	-0.0433
12	-0.0877	6.0530	0.9134	12	-0.1155
13	0.0656	6.6709	0.9183	13	0.0533
14	0.0803	7.6049	0.9089	14	0.0587
15	-0.0032	7.6064	0.9386	15	-0.0170
16	0.0148	7.6388	0.9589	16	0.0475
17	-0.1011	9.1598	0.9351	17	-0.0794
18	-0.1001	10.6660	0.9080	18	-0.1171
19	-0.0271	10.7775	0.9311	19	-0.0255
20	0.0672	11.4691	0.9331	20	0.0495
21	-0.0210	11.5373	0.9513	21	-0.0677
22	0.0026	11.5383	0.9662	22	0.0315
23	-0.0228	11.6204	0.9759	23	0.0002
24	0.0824	12.7014	0.9708	24	0.0762
25	0.0357	12.9057	0.9776	25	0.0942

Appendix D: Air RPM, Pred Formula Air RPM, and ARIMA Forecast vs. Month