Bureau of Transportation Statistics
Technical Report

May 2008

Uncovering Trends in Seasonal Transportation Data
New tool analyzes airline revenue passenger-mile trend data

by Gary Feuerberg, Ph.D.

Transportation activity often varies appreciably from month to month and season to season, making it difficult to discern from the data whether the overall trend for the activity is increasing, unchanging, or decreasing. Traditionally, analysts adjust data to compensate for these fluctuations; however, conventional methods of "controlling" seasonal factors can be limiting—particularly when viewing trends over many years or months. But by *eliminating* the seasonal influence, rather than simply controlling for it, decades of data can be grasped at a glance. In October 2004, the Bureau of Transportation Statistics (BTS) began to remove the seasonal fluctuation in the data used in its monthly Transportation Services Index (TSI)[1]. And, although the approach for removing seasonal factors from airline revenue passenger-miles (RPMs) is examined here, this same approach can be used for most data series that show strong seasonal variation.

Two recent events caused significant declines in airline RPMs—the terrorist attacks of September 11th, 2001, and the SARS (Severe Acute Respiratory Syndrome) outbreak in Asia that occurred in late 2002 and early 2003. The effects of 9/11 are readily apparent to the casual observer even in the raw data, but the impact of the SARS scare on the data is masked by seasonal fluctuations. Although obvious at the time to the airlines that lost revenue and to the passengers who cancelled flights, the impact of the SARS epidemic on RPMs only becomes clear in the data once the seasonal influence is removed.

Typically, holiday and winter recreational travel steadily declines after the first of the year. So, the decline in traffic that occurred from January though April 2003 was not unexpected. But when the seasonal effects are removed from the data, the underlying decline, coincident with the SARS epidemic, is revealed—the number of RPMs was lower than expected.

The method usually employed to investigate airline data trends is to compare data for a particular month from one year to the next. For example, the March 2007 RPM value can be compared with the March 2006 value, or with that same month going back any number of years for which data are available. In this way the seasonal factor is "controlled" and comparisons can be made year by year by matching up the same months. Using this traditional method, one must summarize the changes for each of the 12 months of the calendar and then synthesize the monthly results in order to generalize to the year. However, this method is clumsy when viewing trends occurring over many months or years over the period examined.

The technique BTS adopted to remove distracting seasonal factors in the airline data can be applied to other transportation modes. This report describes BTS' first use of the tool and its potential for supporting analysis.

Adjusting RPM Data To Remove Seasonality

RPMs are one indicator of how the airline industry is performing. Like many transportation variables, RPMs are subject to sizeable fluctuations. No doubt the seasons and holidays have a strong effect on RPMs and can explain most of the variance. But unless measures are undertaken to either control or eliminate the seasonal factors, any underlying trends in the airline data won't become apparent. To clearly see the trends in RPMs, the data must be adjusted to remove seasonal effects. This is called deseasonalizing the data.

The method that BTS employs to adjust the time series data components of the TSI is called X-12 ARIMA, Release 0.2, and is discussed in box A.

Figure 1 displays the raw aviation passenger data that have not been adjusted. The consistent fluctuations that repeat each year in the unadjusted line of raw data are a strong indication that the data are highly seasonal.

---

[1] The Transportation Services Index (TSI) combines available data on freight traffic and passenger travel, drawing on the various modes: air, trucking, rail, waterborne, transit, and pipeline. The data from 10 transportation services are each seasonally adjusted and combined to yield a monthly measure of transportation services output. See www.bts.gov for a full explanation.

Box A: Seasonal Adjustment Methodology

The method discussed in this report to deseasonalize and subsequently smooth time series data has evolved over many decades starting in the 1950s at the U.S. Bureau of Census. Called X-12 ARIMA, Release 0.2, this procedure recently evolved out of the "X-11 Variant of the Census Method II Seasonal Adjustment Program." The name, X-12, means there were 11 variations preceding it. The program was automated by 1965 and is used around the world along with several other procedures. As the most widely used method in seasonal adjustment, BTS adopted it for the TSI based on its acceptance by time series analysts and forecasters.

Fundamentally, X-12 (without ARIMA) is a robust nonparametric[1] method that achieves its estimates through a series of iterative steps. As such, it is an "empirical" approach, as distinguished from a "model-based" approach, to seasonal adjustment. Statistical properties of the estimators, such as confidence intervals, are lacking.

ARIMA modeling was added in 1975. ARIMA, which stands for Auto-Regressive Integrated Moving Average, was introduced to model the time series data in order to produce forecasts. These forecasts are then used to extend the data series before it undergoes adjustments and smoothings. The subsequent seasonal adjustments are found to be more reliable.**[2]**

[1] Nonparametric techniques refer to a branch of applied statistics in which no assumptions are made regarding the underlying form of the data, i.e., "distribution free."

[2] Dominique Ladiray and Benoit Quenneville (2001), *Seasonal Adjustment with the X-11 Method*, Springer-Verlag, New York, p. 9.

Decline of RPMs During the 2nd Quarter 2003

After the aviation RPM data are adjusted to take out the seasonal fluctuation, possible trends are easier to find. Figure 2 shows that the aviation RPM value is generally increasing, suggesting that people are traveling more or further. But a closer look reveals that at least two extended periods of decline occurred within the general tendency of increasing RPMs.

In order to take out all—or nearly all—of the RPM fluctuation, one can use X-12 ARIMA to smooth the deseasonalized data, using the moving average[2] of contiguous data points. The default in X-12 ARIMA for the moving average is 13 months, but it can be set to a different number of months where appropriate.

Note that the smoothed line cannot be precisely defined and graphed. There is no correct "solution" or definitive formula for smoothing data. With more data points, the smoothing will usually "hide" more fluctuation in the data and smooth it out more. On the other hand, with a shorter moving average, the smoothed line will appear jagged and display the smaller changes in the data.

Figure 3 shows the deseasonalized data after it has been smoothed and linked by a line that makes it easier to see trends in the data. Instead of using the default of a mov-ing average of 13 data points, this line uses 5 data points to better pinpoint where the changes lie. Declining periods emerge around the time of the attack of 9/11 in the year 2001 and around the second quarter of 2003, from December 2002 to April 2003. During these four months, there is a continuous drop in the deseasonalized RPM data. Recovery begins in May 2003. Other abrupt changes in the deseasonalized data are shorter in length and can be attributed to the "irregular" component of the model.

The first decline, which is the more dramatic, was apparent even in the raw data (see figure 1) and was expected as a result of the widespread public fear of flying after the 9/11 attacks. But the second decline, the SARS decline, was masked when viewing the raw data because of the seasonal fluctuations discussed below.

Analysis of Regional Routes

The BTS database containing aviation data, TranStats,[3] provides six distinct codes under "region": Atlantic, Domestic, International, Latin America, Pacific, and System. Airlines generally fly in one or two of these regions. This analysis will focus on Pacific region traffic and compare it with traffic in the other regions. But "region" is a term developed by the commercial airlines and lacks scientific precision. The region names are not mutually exclusive and some traffic can logically fall into more than one category. This problem is particularly evident in the minor category

[2] Henderson moving averages are used in X-12 to extract the trend. See Dominique Ladiray and Benoit Quenneville (2001), *Seasonal Adjustment with the X-11 Method*, Springer-Verlag, New York, pp. 36–37.

[3] TranStats is a BTS database found at: http://www.transtats.bts.gov.

Figure 1: Aviation Revenue Passenger-Miles, Raw Data, 1994–2006
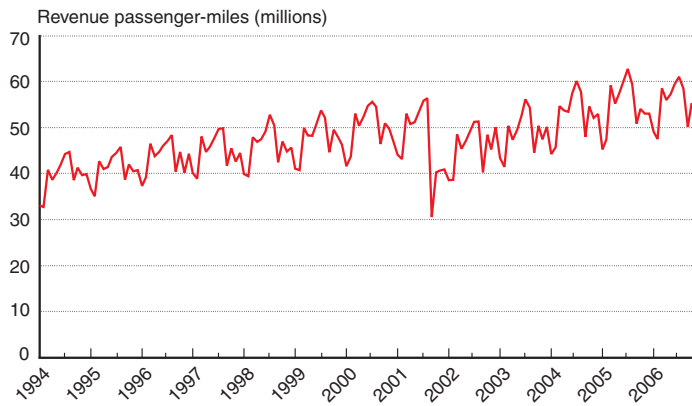
Revenue passenger-miles (millions)

NOTE: RPM = Revenue Passenger-Miles

SOURCE: U.S. Department of Transportation, Research and Innovative Technology Administration, Bureau of Transportation Statistics, available at http://www.transtats.bts.gov/
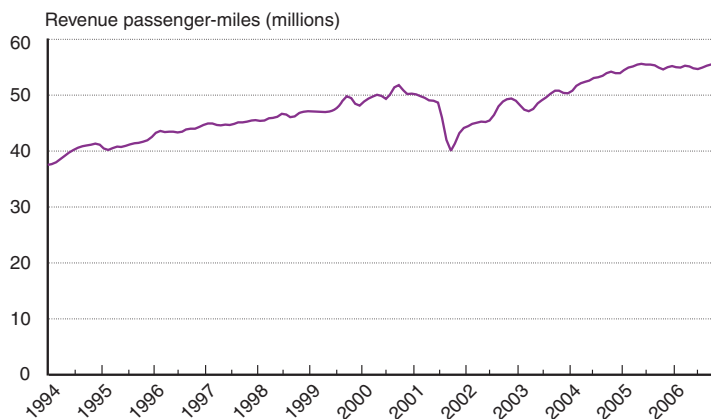
Figure 2: Aviation Revenue Passenger-Miles, Seasonal Influences Removed, 1994–2006
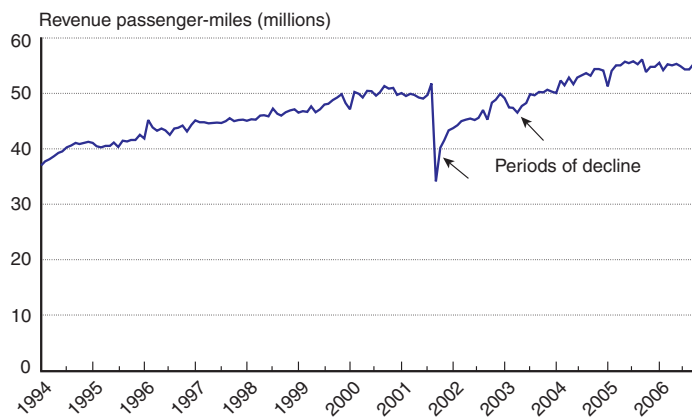
Revenue passenger-miles (millions)

Periods of decline

NOTE: RPM = Revenue Passenger-Miles

SOURCE: U.S. Department of Transportation, Research and Innovative Technology Administration, Bureau of Transportation Statistics, available at http://www.transtats.bts.gov/

Figure 3: "Smoothed" Airline RPM Data Showing Two Distinct Declines: September to November 2001 and December 2002 to April 2003

Revenue passenger-miles (millions)

NOTE: RPM = Revenue Passenger-Miles

SOURCE: U.S. Department of Transportation, Research and Innovative Technology Administration, Bureau of Transportation Statistics, available at http://www.transtats.bts.gov/

of region, International.[4] However, for the most part, the region categories are accurately descriptive. Note that the sixth code, "system," is used very infrequently and is not applicable to this analysis.

From December 2002 to April 2003, the 42.5 percent decline in the Pacific region (see table 1, right column) was quite dramatic compared to declines in the Atlantic, Domestic, and Latin America regions. The overall decline of all regions during this period was 8.4 percent, which becomes the basis to compare traffic changes for each region. With the exception of the residual category of "international," the other regional declines were within about 4 percent of the overall decline, and thus much smaller than Pacific's 42.4 percent decline.

Individual Airlines Analyzed

The BTS database on individual airlines was analyzed in two ways. One method examined the trend of all the airlines, de-seasonalized, while the other method focused only on airlines that declined and determined whether this pattern manifested for travel in the Pacific region.

1. Time series analysis of the RPM trend of the individual airline.

Of the approximately 120 airlines reporting data in the United States in 2003, the top 18 airlines generated three-fourths of total RPMs[5]. These airlines' monthly RPM data were deseasonalized, except for Comair, which was missing too much data. The aim was to see whether a decline in the period December 2002 to April 2003 occurs after the seasonality is removed. Table 2 shows the results.

A trend line of seasonally adjusted data for each airline found that eight of the top airlines showed an unmistak-

able decline in the December 2002 to April 2003 period. Another eight showed no decline. The trend for American Eagle Airlines was unclear, and Comair reported incomplete data[6].

Of the eight major airlines that exhibited the pattern of decline in the 2nd quarter of 2003 (table 2, column 1), only America West Airlines and US Airways did not fly in the Pacific region. Moreover, of the airlines that showed no decline, none flew in the Pacific region (table 2, column 2). Hence, whether an airline flew in the Pacific region is a good predictor in the deseasonalized data of when the airline RPMs declined during the second quarter 2003.

2. Comparison of RPM totals the year before and the year after the "decline" period.

Carefully examining the raw data also supports the trends found in the deseasonalized data, which are easier to detect after the data undergoes deseasonalization. The "Pacific" group of six airlines discussed above exhibited a pattern of decline when comparing the same period one year before (December 2001 to April 2002), and then a recovery one year after the period (December 2003 to April 2004). See table 3. Continental Micronesia Airlines, a smaller airline not in the top 18 but affiliated with Continental, flies frequently in the Pacific region—more than Delta or Hawaiian. It showed the same pattern of decline and recovery during this period.

Continental, Delta Airlines, United Airlines, and Continental Micronesia fit this pattern of decline and recovery. Hawaiian plummeted from 61 million RPMs during the period one year before to 22 million RPMs. It failed to recover by the following year, however. There was very little change in RPMs for Northwest Airlines. Only American Airlines contradicted the pattern.

This pattern of decline and recovery only fit the Pacific region for these particular airlines and not any other region.

---

[4] International refers to a residual region category for traffic to destinations outside the United States reported by carriers that are not required to designate Atlantic, Pacific, or Latin America entities. Not all carriers are assigned Atlantic, Pacific, or Latin America entities, especially when they do not have scheduled flights in those regions.
[5] The number one airline for RPMs in 2002 was American Airlines at 121 billion revenue passenger-miles. The 18th carrier was Frontier with 3.4 billion revenue passenger-miles.

[6] These airline submissions have occasional missing data. In the case of Comair, the amount of missing data precluded further analysis. Three airlines (Frontier Airlines, American Eagle, and Hawaiian) were missing one data point and estimates had to be interpolated.

Table 1: RPMs by Carrier Region, December 2002 to April 2003
(in millions)

Region	December 2002/2003	January 2002/2003	February 2002/2003	March 2002/2003	April 2002/2003	Percent change 12/2002 to 04/2003
Atlantic	6,007	5,481	4,717	5,784	5,465	-9.0
Domestic	42,410	37,206	35,101	42,911	40,371	-4.8
International	513	673	694	834	580	13.1
Latin America	3,355	3,365	2,885	3,348	2,935	-12.5
Pacific	**4,372**	**4,472**	**3,647**	**3,898**	**2,513**	**-42.5**
Total	56,658	51,196	47,044	56,776	51,865	-8.4

NOTES: Totals do not add due to rounding. RPMs = Revenue Passenger-Miles.

SOURCE: U.S. Department of Transportation, Research and Innovative Technology Administration, Bureau of Transportation Statistics, based on data available at http://www.transtats.bts.gov/

**Table 2: Airlines With and Without Decline in RPMs,
December 2002 to April 2003**

Declined	No decline	Indefinite results/missing data
American	Air Tran	American Eagle
America West	Alaska	Comair
Continental	ATA	
Delta	Express Jet	
Hawaiian	Frontier	
Northwest	Jet Blue	
United	Southwest	
US Airway	Spirit Air	

NOTE: RPM = Revenue Passenger-Miles

SOURCE: U.S. Department of Transportation, Research and Innovative Technology Administration, Bureau of Transportation Statistics, available at http://www.transtats.bts.gov/

Table 3: Comparison of Pacific Region RPM Totals (in millions)

	Delta	Northwest	United	American	Continental	Hawaiian	Continental Micronesia
Before (Dec. 2001 to Apr. 2002)	424	7,581	8,830	1,102	1,011	61	1,014
Study period (Dec. 2002 to Apr. 2003)	360	7,700	7,837	1,208	914	22	831
After (Dec. 2003 to Apr. 2004)	453	7,593	9,218	1,552	1,253	19	1,055

NOTE: RPM = Revenue Passenger-Miles

SOURCE: U.S. Department of Transportation, Research and Innovative Technology Administration, Bureau of Transportation Statistics, available at http://www.transtats.bts.gov/

SARS Scares Travelers Away

The few airlines that regularly flew to the Pacific were also the airlines with RPM declines during the 2nd quarter of 2003. The RPM decline suffered by the few airlines serving the Pacific region coincided with news of the SARS outbreak in China, establishing a likely causal connection. The connection with SARS was supported by the lack of a drop in other regions, except for a smaller decline in the Atlantic region.

Travelers to Asia were apparently scared away by the outbreak. The World Health Organization (WHO)[7] reported in October 2004:

"The first known cases of SARS occurred in Guangdong province, China, in November 2002 and WHO reported that the last human chain of transmission of SARS in that epidemic had been broken on 5 July 2003."

and

*"Severe Acute Respiratory Syndrome (SARS) was first recognized as a global threat in mid-March 2003….
By July 2003, the international spread of [SARS] resulted in 8,098 SARS cases in 26 countries, with 774 deaths."*

According to the WHO report, the SARS news discouraged some travelers from flying to Asia. The report included the following statement about the effect that the SARS scare had on travel:

"The epidemic caused significant social and economic disruption in areas with sustained local transmission of SARS and on the travel industry internationally in addition to the impact on health services directly."

It was only by applying deseasonalization to the airline data that the SARS impact became clear. The trend patterns in the data came into full view when the fluctuating seasonal factors were removed, showing the RPM decline during the second quarter of 2003.

Potential for Further Application

Data deseasonalization has great potential for improving analysis of transportation time series data. The approaches outlined in this paper for seasonal adjustment, the application of X-12 ARIMA to airline data, can be used for other transportation data as well. BTS is trying the technique with transportation data pertaining to pipelines, transit, passenger trains, waterborne freight, and air freight. ☯

[7] Guidelines for a Surveillance of Severe Acute Respiratory Syndrome (SARS), October 2004.