MONITORING TRANSPORTATION INDICATORS, AND AN ANALYSIS OF THE EFFECTS OF SEPTEMBER 11, 2001

Peg Young, Bureau of Transportation Statistics, U.S. Department of Transportation, 400 Seventh Street, SW, Room 3430, Washington, DC 20590, USA, peg.young@bts.gov, phone: 202/366-2483, fax: 202/366-2993

Keith Ord, The McDonough School of Business, Georgetown University, Washington, DC 20057, USA, ordk@georgetown.edu, phone: 202/687-2671, fax: 202/687-4031

Topic Area: Applied Statistics

Keywords: forecasting, interventions, transportation

MONITORING TRANSPORTATION INDICATORS, AND AN ANALYSIS OF THE EFFECTS OF SEPTEMBER 11, 2001

Peg Young, Bureau of Transportation Statistics, U.S. Department of Transportation
Keith Ord, The McDonough School of Business, Georgetown University

ABSTRACT:

The Bureau of Transportation Statistics produces a monthly report, called *Transportation Indicators*, which reports on key measures related to the transportation enterprise. The purpose of the report is to provide an up-to-date picture of the transportation sector and a basis for early identification of emerging trends or unusual activity. We describe a procedure that decomposes the time series of interest and generates short-term forecasts for these indicators. We then develop a procedure to compare the new values of these measures to the one-step-ahead forecasts in order to identify those measures that deviated more than expected. We illustrate the performance of this statistical process control procedure on a series describing the percentage of late arriving passenger flights. Finally, we examine the effects of the events of September 11, 2001 on several air transportation series.

1. BTS Transportation Indicators Report

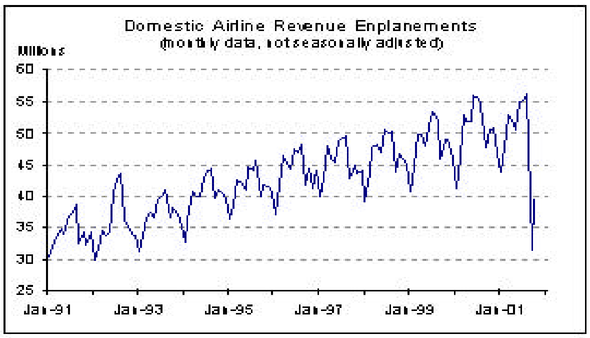
In November 1999, the Bureau of Transportation Statistics (BTS) undertook a project to create a monthly report on key indicators of the transportation system in the United States. The general spectrum of these indicators covers the strategic goals of the Department: Safety, Mobility, Economic Growth, Human and Natural Environment, and National Security. In this paper, we use the term *indicator* to refer to one or more series that relate to a particular topic. For example, under the strategic goal of *Safety*, one indicator refers to the *Transportation of Hazardous Materials*. This indicator includes five series, which give the total number of safety incidents per month, as well as a breakdown by mode for each of air, rail, road and waterborne traffic. As the report is monthly, the data series selected are mostly weekly, monthly and quarterly. Yearly data sets are included only in those instances where more frequent data do not exist. This report is a continually changing document, with new variables being introduced and old variables that prove to be of little value being removed. The first issue of the Transportation Indicators (TI) report, which came out in May 2000, contained over 70 indicators and encompassed over 120 data series; currently, 117 indicators are represented by over 320 time series. An example of an indicator page in the report is provided in Figure 1.

2. Goals of the Monitoring System

The report serves as a resource of up-to-date information on the transportation enterprise that transportation executives could not obtain, in a single source, anywhere else inside or outside the Department. The indicators are offered in a simple form, incorporating a short paragraph describing the data set, a graph of each series over the past 10 years or so, and a table comparing the most recent values of the data series. For highly seasonal data, data comparisons are provided for the same period in the previous year.

TRANSPORTATION INDICATORS | STRATEGIC GOAL: MOBILITY

ENPLANEMENTS ON DOMESTIC FLIGHTS



Revenue enplanements, the number of passengers boarding aircraft, indicate the demand for gate and luggage services. Enplanements differ from the number of trips because passengers may board more than one flight between their origination point and ultimate destination.

Domestic Passenger Aviation	Oct-00	Oct-01
Revenue aircraft enplanements (millions)	50.53	39.80
Percent change from same month previous year	2.88	-21.23

NOTES: The current value is compared to the value from the same period in the previous year to account for seasonality.

The data have been adjusted to have a standard 30-day month by multiplying the data for each month by the ratio: 30/(actual days in month).

The data do not include international flights by U.S. domestic carriers or domestic flights by foreign carriers.

The dramatic changes in the September 2001 data reflect the impact of the terrorist attacks on September 11, 2001, on aviation, including several days in which commercial air operations were suspended.

SOURCE: U.S. Department of Transportation, Bureau of Transportation Statistics, *Air Carrier Traffic Statistics Monthly*, October 2001

U.S. Department of Transportation
Bureau of Transportation Statistics
25
February 2002

FIGURE 1. Example of page from Transportation Indicators report (February 2002 issue).

In addition to providing the time series data within the report, the project also has as an agenda item the task of superimposing a monitoring system for each time series within the report. Such a tracking system should provide a monthly alert system for senior management to advise them when certain indictors have behaved in an 'unexpected' manner. In order to state that behavior was unexpected, we need a method to describe expected behavior. That is, we must first develop a model for each series and then forecast at least one period ahead. We must then create a procedure to compare the forecasts to the new observations, and develop a decision rule to determine whether or not the new data behaved as expected. The terminology "expected/unexpected" may be viewed as broadly equivalent to the more familiar terminology in statistical process control [SPC] of being in or out of control.

This task requires a forecasting model that may be updated easily and quickly, and also permits comparison with recent data. We also need models that allow for deseasonalization (or for decomposition of the seasonal component) so that readers of the report are able to see the underlying trend along with the actual series. Finally, the forecasting process must be able to handle interventions in the data.

3. Description of the data series

The indicators may be classified by eight different criteria.

- A count of the number of time series within that particular indicator; this ranges from one to seven individual time series.
- The strategic goal that is represented: Safety, Mobility, Economic Growth, Human and Natural Environment, or National Security.
- The data generation process: sample survey, enumeration of the whole population, or model-based analysis of empirical data.
- Whether the series are drawn from BTS-controlled sources, such as the data from the Office of Airline Information or from other datasets are outside the control of BTS.
- The recording frequency of the series: weekly, monthly, quarterly, and annual.
- The start date of the series, since some of the series are unusually short.
- Whether or not the series are seasonally adjusted, with consequent implications for the form of forecasting model selected.
- Any additional characteristics not yet captured that would affect the forecast model selected (e.g., apparent interventions).

4. Monitoring using time series

One reason for using time series analysis is to break a series down into its core components (trend, seasonal, and irregular) so that we may examine each separately. The basic ideas for monitoring flow from statistical process control (SPC). The use of time series modeling in SPC follows from the seminal work of Alwan and Roberts (1988, 1995).

In SPC, we conventionally distinguish two sources of variation (c.f. Alwan, 2000, pp. 217-220):

- *Common cause variation:* reflects the natural variation inherent in the process, and
- *Special (or assignable) cause variation:* any variation in the process introduced by a recognizable factor [e.g. a worn tool or a poorly trained operative].

In the present context, we are interested in monitoring changes in a phenomenon over time, and the possible types of assignable cause need to be identified more clearly. Thus, it is useful to divide assignable cause variation into five categories, which we may examine by different means:

- *Additive Outlier (AO):* a factor has a short-term temporary impact on the series, which is resolved within a single observational period. The series then returns to its original state. For example, the effects of a blizzard on the construction industry would typically be of this nature.
- *Temporary Change (TC):* a factor has a relatively short-term impact on the series, which returns to its previous state over several time periods. For example, a prolonged strike in an industry will reduce production, which gradually recovers over the next several months.
- *Level-shift (LS):* a factor causes the series to shift to a new level, and it stays at that new level. For example, a change in reporting requirements might change the level of a series, but not otherwise affect the nature of the phenomenon.
- *Seasonal:* the seasonal pattern in the series may change over time. For example, airlines may change their seasonal pricing strategies, which would lead to a shift in travel patterns.
- *Long-term*: over a period of time, changing conditions may lead to fundamental changes in the series of interest. For example, improved engine design might produce improved fuel efficiency ratings for automobiles, but such an effect would be seen only very gradually in an aggregated series on average miles per gallon.

The components approach to time series enables us to search for each of these assignable causes, while making due allowance for common cause variation. We may use both graphical and numerical procedures to identify problems, as summarized in Table 1. Our focus in this paper is on the first three types of intervention.

Assignable cause	**Graphical procedure**	**Numerical procedure**
Additive Outlier (AO)	Plot recursive [one-step-ahead] residuals	Shewhart chart
Temporary Change (TC)	Plot recursive [one-step-and multi-step-ahead] residuals	Shewhart chart or Cusum charts
Level-shift (LS)	Plot recursive [one-step and multi-step-ahead] residuals	Cusum charts
Seasonal	Plot seasonal component	Check variance of seasonal component
Long-Term	Plot trend component	Cusum chart

TABLE 1. Graphical and numerical procedures for the identification of assignable causes in time series.

5. The structural time series model

Although terms such as 'trend' and 'seasonal' are intuitively appealing, they are mental constructs; we cannot observe them directly. Therefore, we use a structural modeling approach that treats them as *unobserved components* (Harvey, 1989; Harvey and Shephard, 1993). We used the STAMP [Structural Time series Analyzer, Modeller and Predictor] software in conjunction with GiveWin; for details, see Koopman et al., (2000).

We define the components at time *t* as follows: trend = μ_t; slope = β_t; seasonal component = γ_t; and irregular component = ε_t. We assume that the process is observed at unit time intervals *(t, t+1,…)* and that there are *s* such intervals in a year. We then allow each component to evolve over time according to the specifications:

$$m_t = m_{t-1} + b_{t-1} + h_t \quad (1)$$

$$b_t = b_{t-1} + V_t \quad (2)$$

and

$$g_t + g_{t-1} + ... + g_{t-s+1} = w_t \quad (3)$$

The quantities η_t, ζ_t, and ω_t represent zero mean, random shifts in the corresponding component. We assume such shifts to be independent of one another and uncorrelated over time; we also assume that they are independent of the 'irregular' component defined below. Equations (1)-(3) are known as the *state* or *transition* equations since they describe the underlying state of the process, or the transition of the components from one time period to the next.

Expressions (1) and (2) provide a very general framework for describing the evolution of the trend. If the process being modeled does not require all these components they can be dropped from the specification. The components are tested in sequential fashion as follows (Harvey, 1989, pp. 248-56):

1T. Does the slope disturbance term have positive variance? [Zero variance corresponds to removing that term.]

2T. Does the level disturbance have positive variance?

3T. If the slope disturbance is dropped, does the slope differ from zero?

If all three tests produced negative outcomes the trend term would be reduced to a constant.

When the time series is seasonal, we check:

1S. Does the seasonal disturbance term have positive variance?

2S. If the seasonal disturbance is dropped, are the seasonal components significantly different from zero? [Is there a seasonal pattern?]

If we drop the disturbance term we are left with a "classical" model with fixed seasonals. If the seasonal pattern is rejected completely, we reduce the model purely to its trend components.

The state of the system is related to the observed series by the *observation (or measurement) equation*:

$$y_t = m_t + g_t + e_t \tag{4}$$

where ε_t denotes the 'irregular' component. The irregular component has zero mean and is assumed to be unrelated to its own past (i.e. not predictable) and independent of the disturbances in the state equations.

Estimation proceeds by maximum likelihood (Harvey, 1989, pp. 125-128). Operational details are provided in Koopman et al. (2000, section 8.3). The key parameters are the four variances corresponding to the disturbance terms [σ_ε^2, σ_η^2, σ_ζ^2 and σ_ω^2]. Note that we assume these variances are constant over time; the time series may need to be transformed to justify this assumption, at least to a reasonable degree of approximation. The four variance terms control the form of the model, allowing each of level, slope and seasonal to be stochastic or fixed; slope and seasonal may be present or absent. Table 2 illustrates the principal variations. If fixed components are included in a model, the corresponding terms appear in the state equations (e.g. fixed seasonal coefficients) but the variance term is zero. If the components are stochastic, the same terms appear in the model, but the variance is strictly positive. The most general form is the Basic Structural Model (BSM), in which all components are stochastic. The BSM forms the starting point for the model development process, and is the standard form employed in STAMP. We then 'tested down' to eliminate any components that were not required for a particular series. An initial set of interventions (prior to September 2001) was identified using the notes provided in the original Transportation Indicators documentation, combined with an initial analysis using the AUTOBOX software (produced by David Reilly of Automatic Forecasting Systems)

Table 2: Some of the principal models in the structural framework

[Based upon Koopman et al. (2000), page 141]

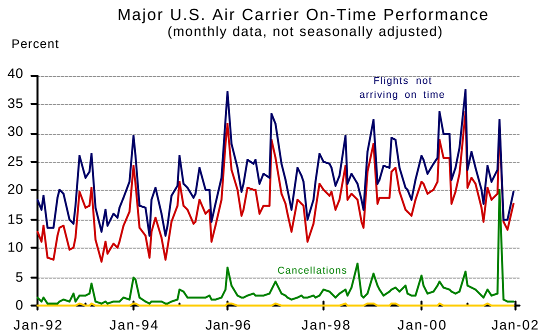
Type of model	σ_ε^2	σ_η^2	σ_ζ^2	σ_ω^2
Level only				
Constant mean	yes	0	0	0

Local level (LL)	yes	yes	0	0
Random walk (RW)	0	yes	0	0
Trend				
Deterministic	yes	0	0	0
Local level with fixed slope	yes	yes	0	0
Random walk with fixed drift	0	yes	0	0
Local linear trend (Holt)	yes	yes	yes	0
Smooth trend	yes	0	yes	0
Second difference	0	0	yes	0
Seasonal (with selected trend)				
Fixed seasonals	(yes/0)	(yes/0)	(yes/0)	0
Varying seasonals	(yes/0)	(yes/0)	(yes/0)	yes
Basic Structural Model (BSM)	yes	yes	yes	yes

6. Analysis of airline delays

To illustrate the proposed forecasting and monitoring techniques with respect to the impact of September 2001, we consider an example that has gained considerable publicity of late – airline delays. Figure 2, a page drawn from the February 2002 issue of the TI report, provides a summary of the on-time performance measures for the major US carriers for the past 10 years.

MAJOR U.S. AIR CARRIER ON-TIME PERFORMANCE



The number of flights not departing or arriving on time, cancellations, and diversions are measures of service quality.

These indicators are strongly seasonal and are affected by weather and heavy demand in winter and summer months, respectively.

On-Time Performance	Dec-00	Dec-01
Number of scheduled flights	475,398	394,787
Percent change from same month previous year	*1.16*	*-16.96*
Percent of flights not arriving on time	37.59	19.77
Change from same month previous year	*15.59*	*-17.83*
Percent of flights not departing on time	33.80	17.68
Change from same month previous year	*15.38*	*-16.12*
Percent of cancelled flights*	6.00	0.86
Change from same month previous year	*4.26*	*-5.13*
Percent of diverted flights**	0.31	0.14
Change from same month previous year	*0.12*	*-0.17*

* Also counted in flights not arriving or departing on time.
** Also counted in flights not arriving on time.

NOTES: The current value is compared to the value from the same period in the previous year to account for seasonality.

The data cover the 10 largest U.S. air carriers. A scheduled operation consists of any nonstop segment of a flight. The term "late" is defined as 15 minutes after the scheduled departure or arrival time. A cancelled flight is one that was not operated but was listed in a carrier's computer reservation system within seven calendar days of the scheduled departure. A diverted flight is one that left from the scheduled departure airport but flew to a destination point other than the scheduled destination point.

Aloha Airlines began reporting in October 2000 and is included here starting in October 2001. For comparability, the year-ago changes and growth rates are based on data that excludes Aloha. American Eagle began reporting in January 2001, will be excluded here until one year's data is available to retain comparability with previous years.

The dramatic changes in the September 2001 data reflect the impact of the terrorist attacks on September 11, 2001, on aviation, including several days in which commercial air operations were suspended.

SOURCE: U.S. Department of Transportation, Bureau of Transportation Statistics, Airline Service Quality Performance data.

FIGURE 2. Major US air carrier on-time performance (February 2002 issue).

Included in this set of measures are

- Flights not arriving on time,
- Flights not departing on time,
- Cancellations, and
- Diversions

For our analysis, we selected the percent of flights not arriving on time as the variable of interest, which we refer to as 'late arrivals.' The percentage figure was used to eliminate the effects of the growth in traffic, although increased traffic without the addition of extra capacity would be a source of delay in itself. A graph of this single data series, starting in September 1987, is provided in Figure 3.

Prior to modeling the data in STAMP, the late arrival data were analyzed in AUTOBOX to find an initial set of interventions. Two significant pulses [AO] within the time period of September 1987 through August 2001 were found: January 1996, and December 2000. These two interventions were incorporated into the STAMP modeling process. Our analysis of the series, using STAMP, revealed that the most appropriate model was one with stochastic level, no slope and fixed seasonals.

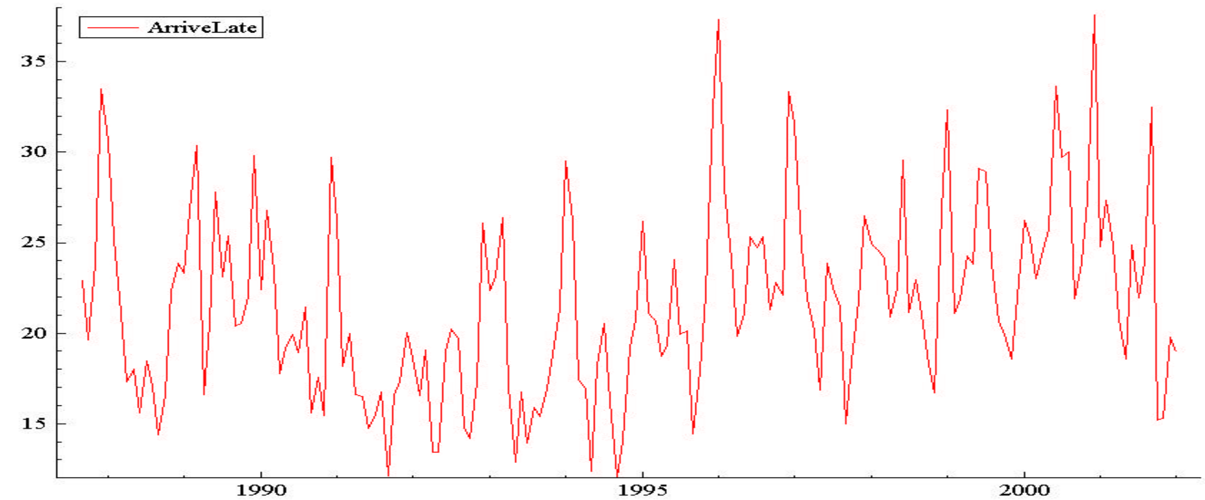


FIGURE 3. Late arrivals as a percent of total operations for major US air carriers.

This model yields the outputs shown in Figures 4-8. Figure 4 shows the smoothed trend and seasonal components; the *smoothed* versions are the better choice for gaining a perspective on the evolution of the series as the estimates use observations both before and after the time period in question.

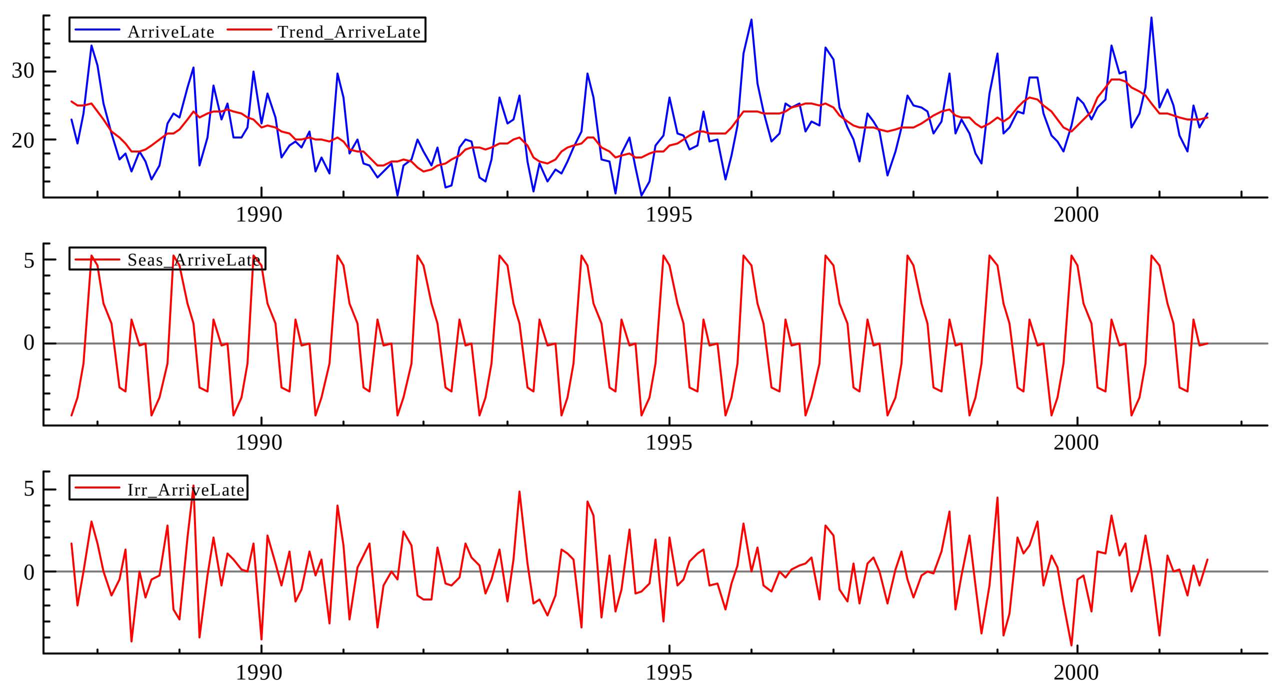


FIGURE 4. Smoothed components of the airline delays series generated by STAMP.

When these plots are compared with the *filtered (or forecast)* components in Figure 5, the increased roughness of the latter set becomes evident. However, the filtered components use only the observations up to the time period under investigation and are therefore more useful for monitoring purposes.

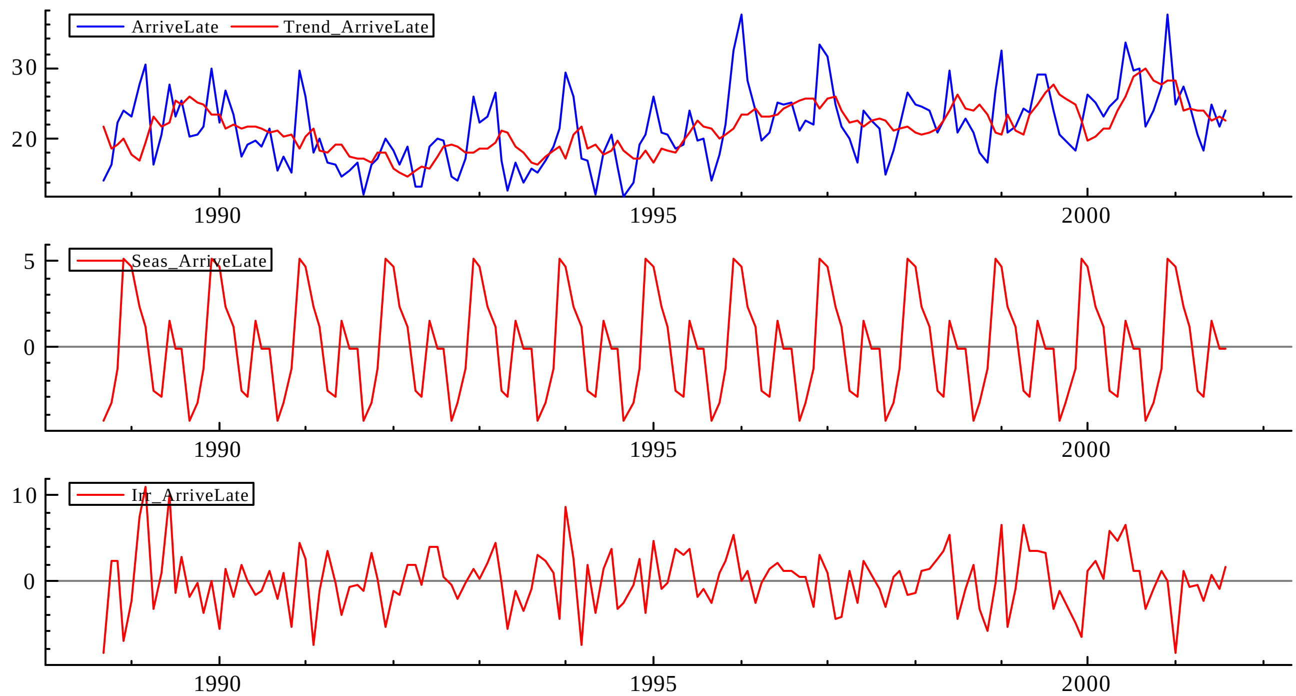


FIGURE 5. Filtered components of the airline delays series generated by STAMP.

Note that the optimal model has been based upon the data from September 1987 through August 2001. Since the model has been specified, the holdout sample of data from September 2001 through January 2002 is placed back into the data set, and same model is fitted to the full set of data. We now analyze the resultant residuals with the Shewart and Cusum charts to check for unusual observations.

Figure 6 shows the standardized residuals for the full fitted series and highlights the impact of post-September 11, 2001. The residuals analysis in the Shewart chart indicates a sharp rise in late arrivals in September, followed by a severe decline in late arrivals in October 2001.

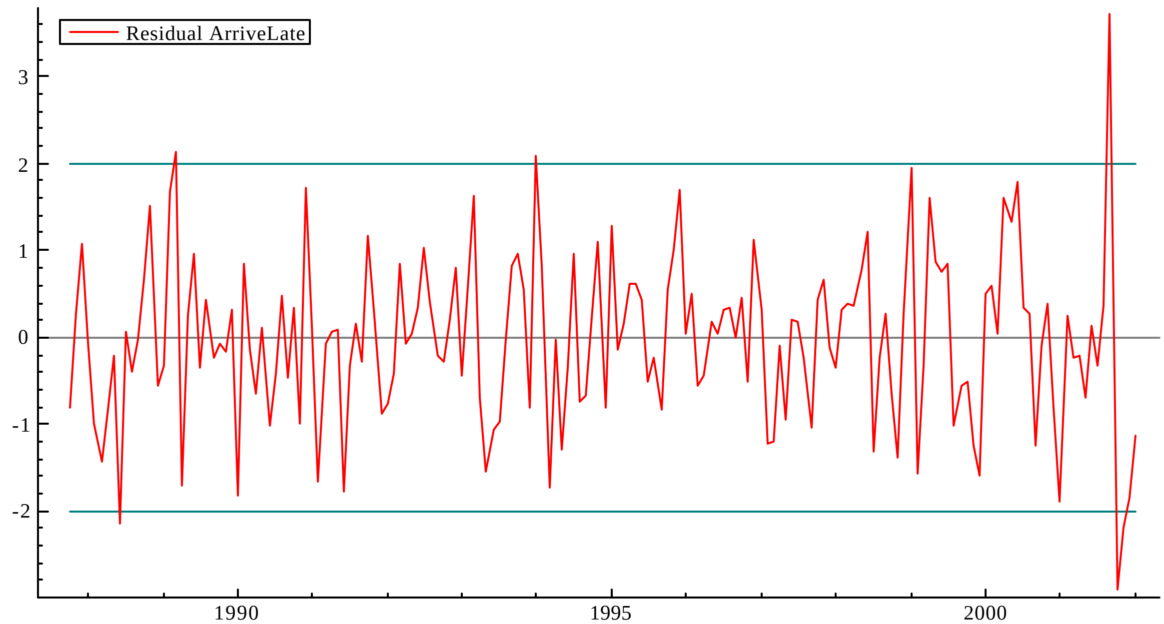


FIGURE 6. Airline delay series: Shewart chart of standardized residuals.

In order to test for long-term assignable causes, we ran a Cusum test on the residuals (see Figure 7).

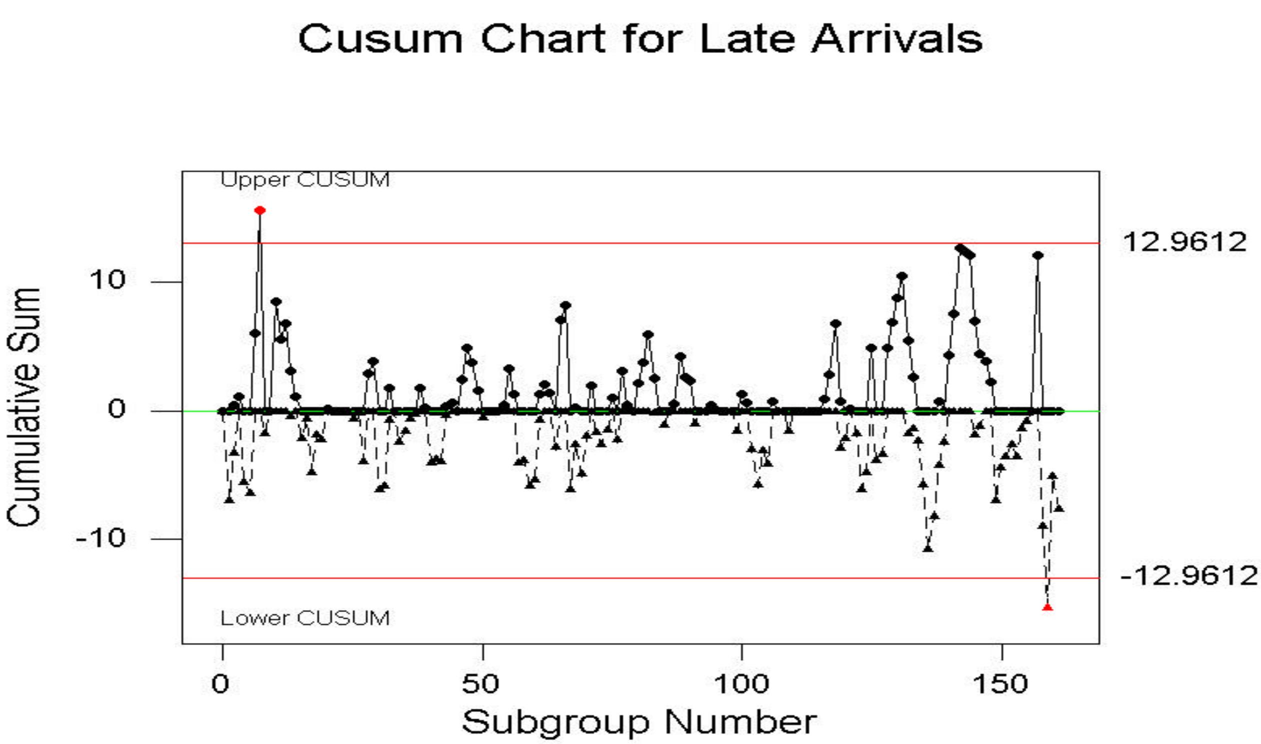


FIGURE 7. Cusum chart for Percent late arrivals, resetting after each alert.

We see an alert in October 2001, which indicates a level shift at this point in time. This concurs with the October drop in the Shewart analysis. The two charts together highlight a pulse in September 2001 and a level shift in October 2001. These interventions are also noted in the STAMP results for the full data set. The final fitted underlying trend for the full set of data is shown in Figure 8. Although the overall performance appears to be satisfactory, the further declines in the trend after October 2001 seem implausible and suggest the need for further modeling. Since data for such a modeling exercise is necessarily very limited, we need to use our insights on likely future developments, as illustrated in the next section.

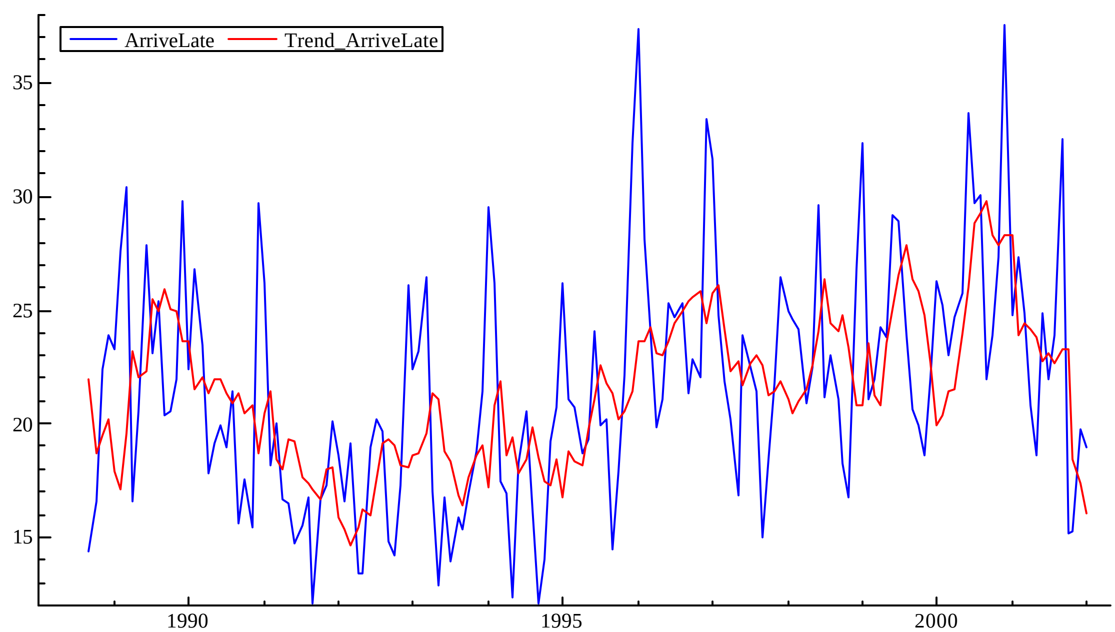


FIGURE 8. Final trend for late arrivals.

6. A General Approach to Intervention model

Since the impact of 9/11 is our primary focus, this section is written in that context. However, a similar approach may be adopted in any circumstances where an unusual event disrupts the phenomenon under study.

We allowed the effects of 9/11 to be represented by three components:

- A purely temporary effect (or additive outlier, AO) relating to the month of September only.
- A permanent effect (or level shift, LS) that changed all mean values of the series from October 2001 on. Note that we could have started this factor in September, but we feel that the present construction affords a simpler interpretation.
- A temporary change (TC) that started in October 2001 and gradually disappeared. Again, we could have started this effect in September, but felt that October provided a simpler interpretation.

A major question with the temporary change is the rate at which the decay takes place. Indeed, the additive outlier may be viewed as a transient change with a dampening or decay factor, d = 0, or sufficiently small to largely disappear within one time period. Likewise, a level shift may be viewed as a temporary change with d = 1. Chen and Liu (1993) recommend d = 0.7 as a convenient choice, which we may motivate by noting that the "half-life" of the TC is then about three months. That is, in months 1, 2, and 3, the weights assigned to the TC are 1.0, 0.7 and 0.49, reducing to about half the starting weight. Likewise, a value of d = 0.8 corresponds to a half-life of about four months and d = 0.9 to a half-life of just over seven months. A value greater than 0.9 becomes almost indistinguishable from a level shift, whereas anything appreciably less than 0.7 may be represented by a one or (at most) two period AO.

We should keep in mind that we have a very limited amount of data with which to estimate these effects, and that the direct estimation of decay rates is difficult even when we have a considerable number of observations after the intervention. Although many interventions are unique in nature, the notion of the time taken to recover from the effect is quite well understood. Thus, subject matter specialists can often provide reasonable estimates of the half-life for a new intervention, even though the magnitude of the effect cannot be reliably assessed in advance.

Based both upon the argument of Chen and Liu (1993) and observations about likely recovery times, we used d = 0.7 in our study for all series. Our procedure was as follows:

1. Develop a model for the series prior to September 2001, incorporating AO and LS outliers where needed (as for late arrivals in section 5).
2. Run the same model with AO, LS and TC components as specified above and "test down" to eliminate insignificant coefficients.

6.1 Some empirical results

To examine the proposed approach, we considered five time series relating to air travel. We selected these series as they are among those showing the most severe effects. Since we had between 4 and 6 observations post 9/11 (from September to December or February depending on the series), it was only to be expected that the LS and TC estimates would be highly correlated, and the TC dropped out in three of the five series. The results are given in Table 2.

Time series	Values *	AO (Sept.)	LS (Oct.)	TC (Oct.)
Cancellations	6	18.1 (.0000)	-1.6 ** (.0001)	n.s
Enplanements	4	-16.4 (.0000)	-4.0 (.0324)	-7.3 (.0005)
Late Arrivals	6	14.1 (.0000)	-6.1 (.0427)	n.s.
Miles Traveled	4	-8338 (.0003)	4000 (.0077)	n.s.
Revenue Passenger Miles	4	-13.9 (.0000)	-3.7 (.0267)	-6.5 (.0006)

* Number of observations in the series after August 2001

** LS starts in November 2001, for improved fit

n.s. Not significant.

Table 2: Results of intervention analyses for five air travel series.

The conclusions that may be drawn from Table 2 are as follows:

- In all cases, large adverse effects were identified in September, as expected.
- Cancellations and Late Arrivals showed negative level shifts, reflecting the reduced amount of traffic in the months following 9/11. As air traffic recovered, we might have expected this effect to disappear, but the early evidence suggests that the changes in operating procedures are maintaining the shift at a lower level.
- The Miles Traveled series shows an increase in level, which is consistent with the data but was not expected and for which we do not have an explanation.
- Enplanements and Revenue Passenger Miles showed slower adjustments than the other series and both suggest a temporary effect that is about 80% greater than the permanent level shift. All these effects were negative, indicating the adverse effect on the airline industry. The slower nature of the adjustment process and the reduced final impact are both worthy of note.
- Although the details are not reproduced here, the diagnostics for each series indicated that the descriptions were consistent with the data available. Since the same set of three interventions was

applied to each series, this provides some evidence that the descriptions are reasonable, although further data are clearly needed to validate that claim.

Final Comments

As the number of data points available increases, we can expect more reliable estimates of the impact of the 9/11 attack. Our intention is to preserve these current estimates and then compare them with later estimates when a clearer pattern has emerged. Also, we plan to try the approach on other series. However, we are cautiously optimistic that the proposed approach offers a way forward in dealing with interventions at the end of a series.

References

Alwan, L. C. (2000). *Statistical Process Analysis*. Irwin McGraw-Hill: Boston.

Alwan, L.C. and Roberts, H.V. (1988). Time-series modeling for statistical process control. *Journal of Business and Economic Statistics,* 16, 83-95.

Alwan, L.C. and Roberts, H.V. (1995). The problem of misplaced control limits. *Applied Statistics,* 44, 269-278.

Harvey, A.C. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter.* Cambridge University Press: Cambridge, UK.

Koopman, S. J., Harvey, A.C., Doornik, J.A. and Shephard, N. (2000). *Stamp: Structural Time Series Analyser, Modeller and Predictor*. Timberlake Consultants: London, UK.