Mercury Marine & Automatic Forecasting Systems & Business Modelling Associates

Data Cleansing and Automatic Procedures

Tom Reilly Vice President

MERCURY
#1 On The Water™

Our Company

- 1975 Incorporated
  - First-to-market Forecasting package
- 1976 "AutoBJ" available on Mainframe Time-sharing Services – IDC, CSC and Compuserve
- 1982 Autobox 1.0 launched DOS version on the PC
- 1991 Windows version
- 1996 Batch Version
- 1999 UNIX/AIX version
- 1999 Callable DLL version for "plug and play" into other systems
- 2011 - Partners with South Africa based Business Modelling Associates
www.businessmodellingassociates.com

Why is Autobox's Methodology Different?

- **Automatically creates a customized model for every data set.**

- **Automatically identifies and corrects outliers in the historical data and for the causal variables to keep the model used to forecast unaffected (Pulses, seasonal pulses, level shifts, local time trends)**

- **Automatically will identify and incorporate the time lead and lag relationship between the causal variables the variable being predicted**

- **Automatically will delete older data that behaves in a different "model" than the more recent data (i.e. Parameter Change detection via Chow Test)**

- **Automatically will weight observations based on their variance if there has been changes in historical volatility (i.e. Variance Change detection via Tsay Test)**

- **Automatically will identify intermittent demand data and use a special modelling approach to forecast the lumpy demand**

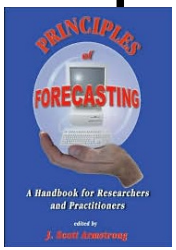
MERCURY MARINE

business modelling ASSOCIATES

AUTOMATIC FORECASTING SYSTEMS

Awards

- Picked as the "Best Dedicated Forecasting" Software in the "Principles of Forecasting" text book (Go to page 671 for overall results)



- Placed 12th in the "NN5" 2008 Forecasting Competition on "Daily data" (Click here to see www.neural-forecasting-competition.com results), but 1st among <u>Automated</u> software.

NN5
www.neural-forecasting-competition.com

- Placed 2nd in the "NN3" 2007 Forecasting Competition on "Monthly data" (Click here to see www.neural-forecasting-competition.com), and 1st on more difficult data sets.

NN3
www.neural-forecasting-competition.com

MERCURY MARINE

business modelling ASSOCIATES

AUTOMATIC FORECASTING SYSTEMS

Mercury Marine

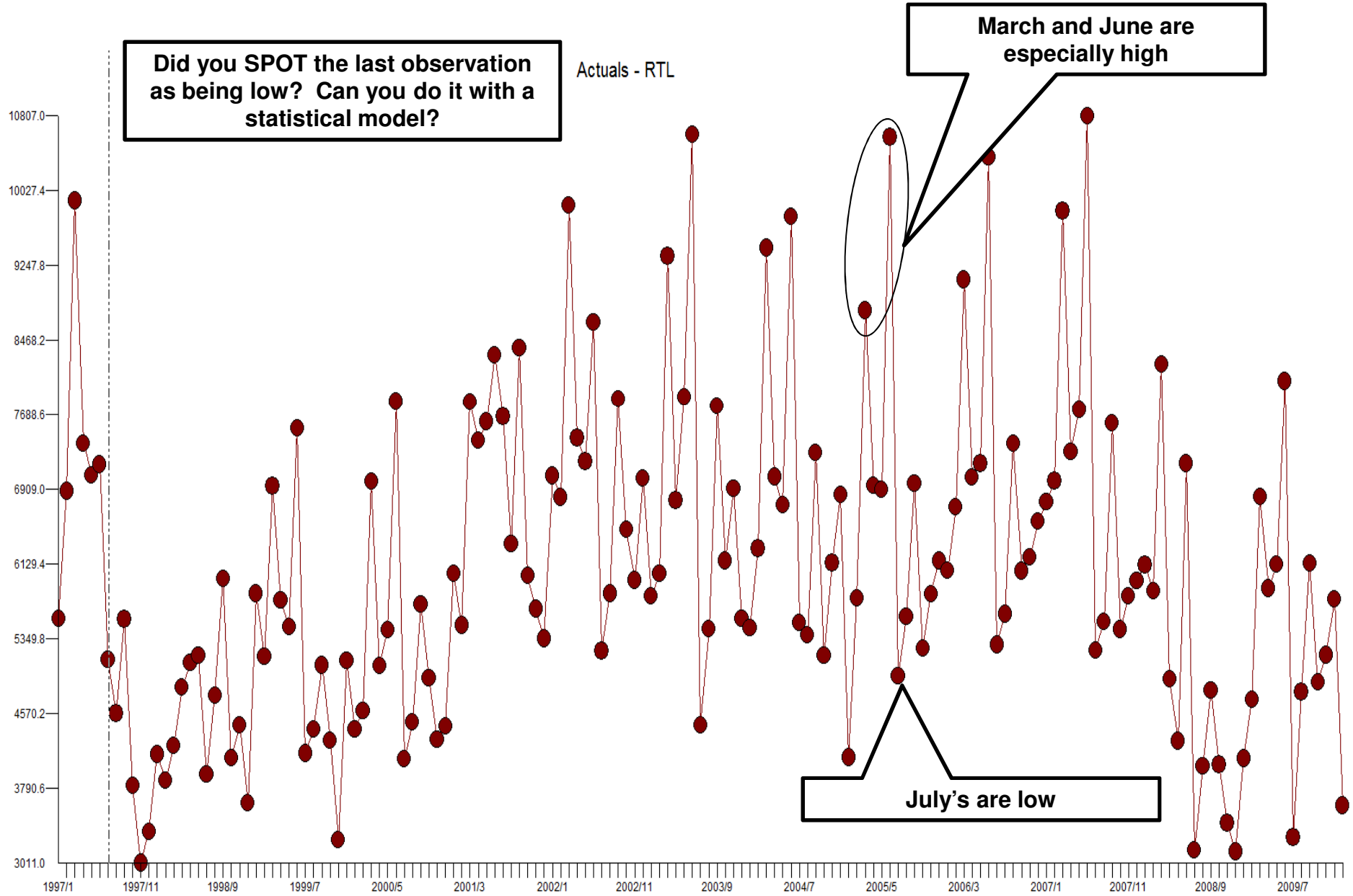
- Mercury Marine's forecasting process needed an improvement from basic approaches that did not provide accurate forecasts.
- They realized that their forecasts were a function of the software and its methodology.
- We undertook an effort to track accuracies in a more rigorous way to understand our service levels
- Mercury Marine published an article in APICS in 2009 with a full discussion of Autobox and its usage http://autobox.com/pdfs/apics_rev8.pdf

MERCURY MARINE

business modelling ASSOCIATES

AUTOMATIC FORECASTING SYSTEMS

Monthly Outboard Motor shipments



MERCURY MARINE

AUTOMATIC FORECASTING SYSTEMS

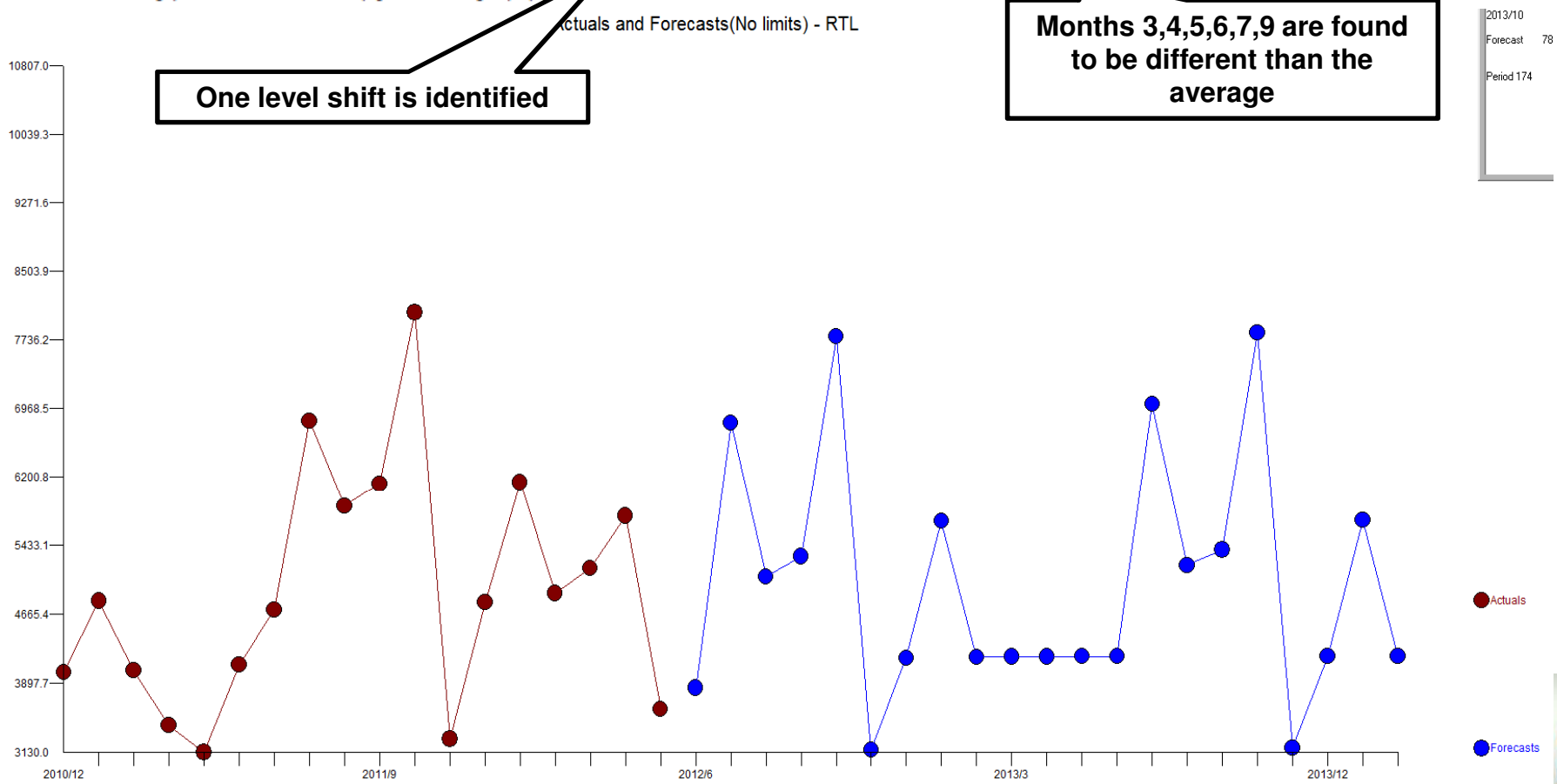
Howzit???? It could be better!

```
Y(T) =  5772.6
      +[X1(T)][(+  2819.4    )]          G_MONTLYEFFECT3
      +[X2(T)][(+  1014.2    )]          G_MONTLYEFFECT4
      +[X3(T)][(+  1185.5    )]          G_MONTLYEFFECT5
      +[X4(T)][(+  3620.3    )]          G_MONTLYEFFECT6
      +[X5(T)][(+  1514.2    )]          G_MONTLYEFFECT9
      +[X6(T)][(-  1014.8    )]          :SEASONAL PULSE  2003/  7          I~S00051RTL
      +[X7(T)][(-  1576.7    )]          :LEVEL SHIFT     2008/  4          I~L00108RTL
     +     [(1-  .597B** 1)]**-1  [A(T)]
```

Seems like a decent model and forecast IF you don't acknowledge that the last value is an anomaly

One level shift is identified

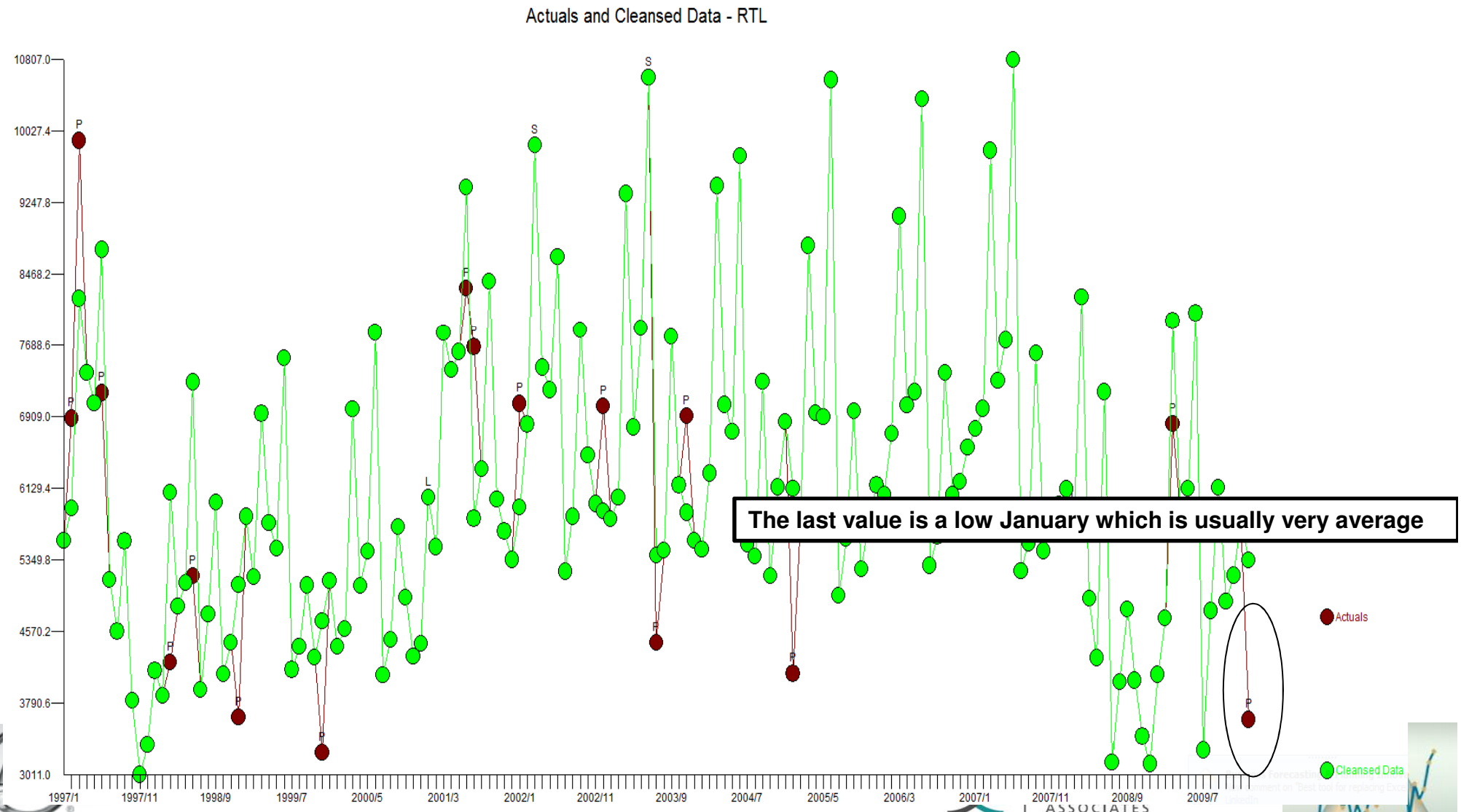
Months 3,4,5,6,7,9 are found to be different than the average



MERCURY MARINE

AUTOMATIC FORECASTING SYSTEMS

Actuals and Cleansed of Outliers Plot



The last value is a low January which is usually very average

MERCURY MARINE

AUTOMATIC FORECASTING SYSTEMS

Model is more robust

```
Y(T) =  4436.9
       +[X1(T)][(+  332.30    )]          G_MONTLYEFFECT2
       +[X2(T)][(+  2414.6    )]          G_MONTLYEFFECT3
       +[X3(T)][(+  1166.8    )]          G_MONTLYEFFECT4
       +[X4(T)][(+  1271.7    )]          G_MONTLYEFFECT5
       +[X5(T)][(+  3140.5    )]          G_MONTLYEFFECT6
       +[X6(T)][(-  475.28    )]          G_MONTLYEFFECT7
       +[X7(T)][(+  1517.3    )]          G_MONTLYEFFECT9
       +[X8(T)][(+  1173.9    )]          :SEASONAL PULSE  2003/  6
       +[X9(T)][(+  1868.3    )]          :PULSE           2001/  7
       +[X10(T)[(-  2112.7    )]          :PULSE           1998/  6          I~P00018RTL
       +[X11(T)[(-  1850.2    )]          :PULSE           1998/  3          I~P00015RTL
       +[X12(T)[(-  2011.8    )]          :PULSE           2005/  1          I~P00097RTL
       +[X13(T)[(-  1570.2    )]          :PULSE           1997/  6          I~P00006RTL
       +[X14(T)[(-  1736.2    )]          :PULSE           2010/  1          I~P00157RTL
       +[X15(T)[(-  1447.4    )]          :PULSE           1998/ 12          I~P00024RTL
       +[X16(T)[(-  1432.3    )]          :PULSE           1999/ 11          I~P00035RTL
       +[X17(T)[(-  1378.8    )]          :LEVEL SHIFT     2008/  2          I~L00134RTL
       +[X18(T)[(+  1462.4    )]          :LEVEL SHIFT     2001/  1          I~L00049RTL
       +[X19(T)[(+  1049.7    )]          :PULSE           2003/ 11          I~P00083RTL
       +[X20(T)[(+  1127.8    )]          :PULSE           2002/  1          I~P00061RTL
       +[X21(T)[(+  1145.4    )]          :PULSE           2002/ 12          I~P00072RTL
       +[X22(T)[(+  [illegible]  )]       :SEASONAL PULSE  2002/  3          I~S00063RTL
       +[X23(T)[(+  [illegible]  )]       :PULSE           2001/  6          I~P00054RTL
       +[X24(T)[(+  [illegible]  )]       :PULSE           2009/  3          I~P00147RTL
       +[X25(T)[(-  952.66    )]          :PULSE           2003/  7          I~P00079RTL
       +[X26(T)[(+  1720.5    )]          :PULSE           1997/  3          I~P00003RTL
       +[X27(T)[(+  975.84    )]          :PULSE           1997/  2          I~P00002RTL
      +     [(1-  .666B** 1)]**-1  [A(T)]
```

Months 2 is now identified to be unusual

Low January in most recent observation and one of many outliers identified

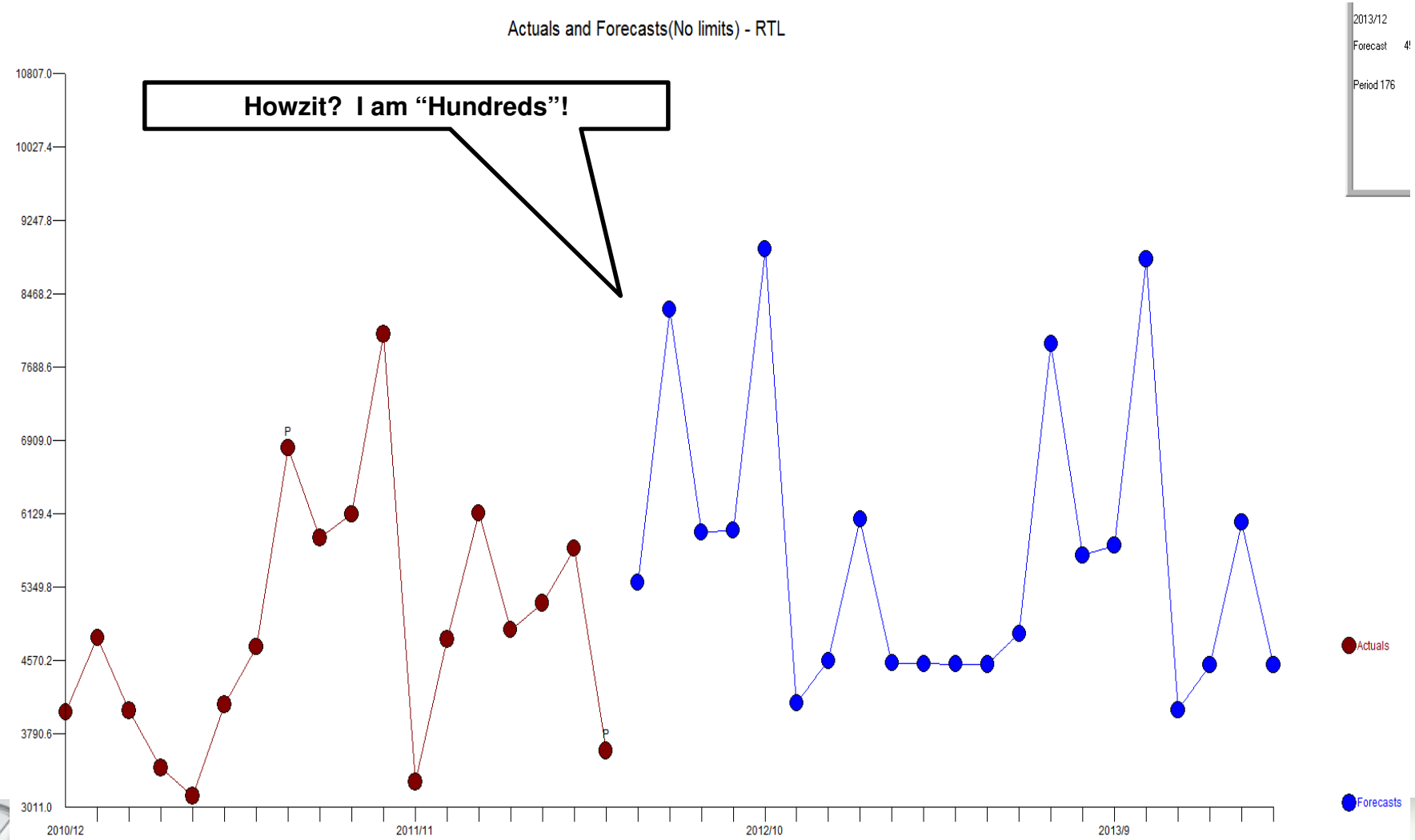
Two level shifts are now identified

MERCURY MARINE

business modelling ASSOCIATES

AUTOMATIC FORECASTING SYSTEMS

Level of the forecast no longer being skewed due to outlier

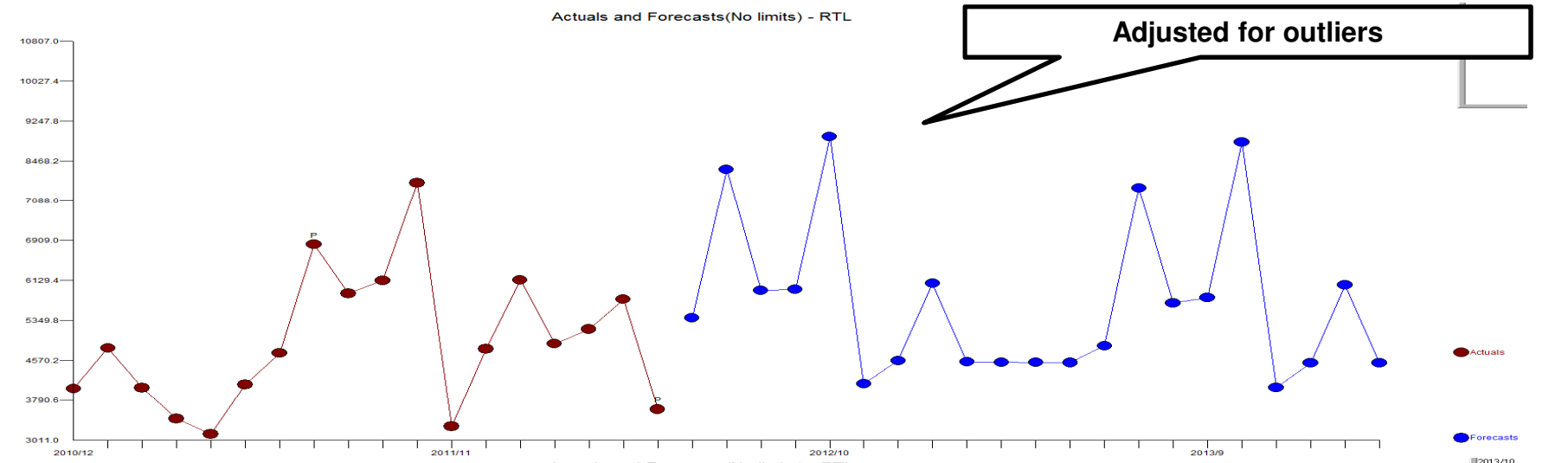


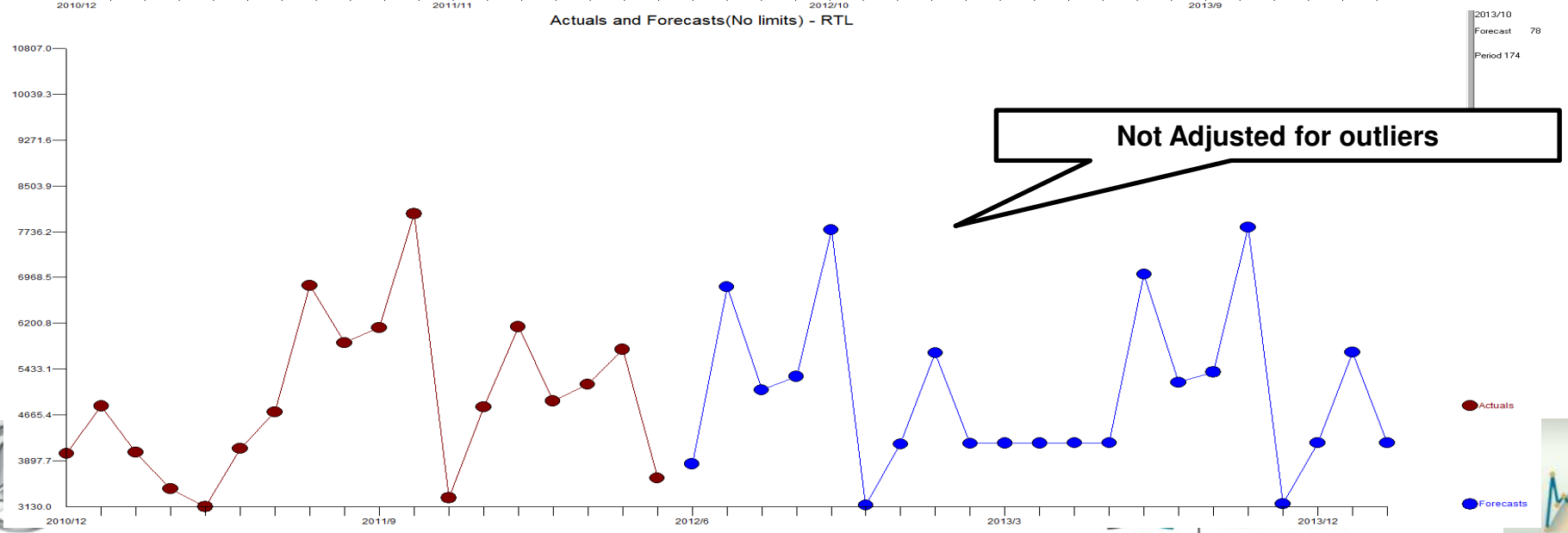
MERCURY MARINE

business modelling ASSOCIATES

AUTOMATIC FORECASTING SYSTEMS

If you ignore outliers you get a bad forecast





MERCURY MARINE

AUTOMATIC FORECASTING SYSTEMS

AMR/Gartner Tells us that the S&OP Process doesn't think that Technology is relevant

- 60% Change Management
- 30% Process
- 10% Technology
  - Source – 2010 Supply Chain Executive Conference – Jane Barrett & Noha Tohamy http://www.gartner.com/it/content/1393400/1393424/august_18_supply_chain_jbarrett.pdf
- Consider where the message is coming from -They are selling consulting and not technology!
- Also consider that the world has written off any value in a quality statistical forecast as typical software does it so poorly that they shift the focus on the S&OP process to "fix" the BAD baseline forecast and lean on qualitative adjustments instead of quantitative adjustments to future events.

MERCURY MARINE

business modelling ASSOCIATES

AUTOMATIC FORECASTING SYSTEMS

Data Cleansing

Researchers Lead The Way ! Some Developers Follow !

- Ruey Tsay "Outliers, Level Shifts, and Variance Changes in Time Series" *Journal of Forecasting* 1988, *7*(1), pp. 1.
- Ruey Tsay "Time Series Model Specification in the Presence of Outliers" *Journal of the American Statistical Association*, 1986, *81*(393), pp. 132-41
- Bell, W. (1983). "A Computer Program for Detecting Outliers in Time Series," in American Statistical Association 1983 Proceedings of the Business Economic Statistics Section, Toronto, pp. 624-639.
- Chang, I., and Tiao, G.C. (1983). "Estimation of Time Series Parameters in the Presence of Outliers," Technical Report #8, Statistics Research Center, Graduate School of Business, University of Chicago, Chicago.
- Chen, C. and G. C. Tiao (1990) "Random Level Shift Time Series Models, ARIMA approximations, and Level Shift Detection" The Journal of Business and Economic Statistics , January, 1990, p.81-96.
- Box, G.E.P., and Tiao, G. (1975). "Intervention Analysis with Applications to Economic and Environmental Problems," Journal of the American Statistical Association, Vol 70, pp. 70-79.

What is unusual?

- We are asked as children "What doesn't belong?" We build a "model" in our head as to what is usual and what is unusual.

A	B
C	71

- Early researchers thought that all unusual values can be detected when they occur outside some pre-set range such as +- 3 sigma around the mean. Typically, data was plotted and a visual review was to identify anomalies

- Typically, statistics like the standard deviation were calculated and 3 sigma bands were put around the mean to identify outliers. The reality is that the standard deviation that was calculated is skewed upwards by the outlier itself so this approach is not reliable. An assumed model was used in this process where the data was subtracted by the mean to get residuals. Who is to say that the mean is in fact the correct model for the data?

What is unusual?

- When data needs to be cleansed this suggests that we have omitted an important variable in the modeling process. This omitted deterministic(0 or 1 dummy variable) variable may be either known to us or unknown to us. Detecting this phenomenon often leads directly to “hypothesis generation” where data suggests theory, such as the need for an omitted event.

- Care must be taken not to falsely identify anomalies that are systematic such as a seasonal pulse variable.

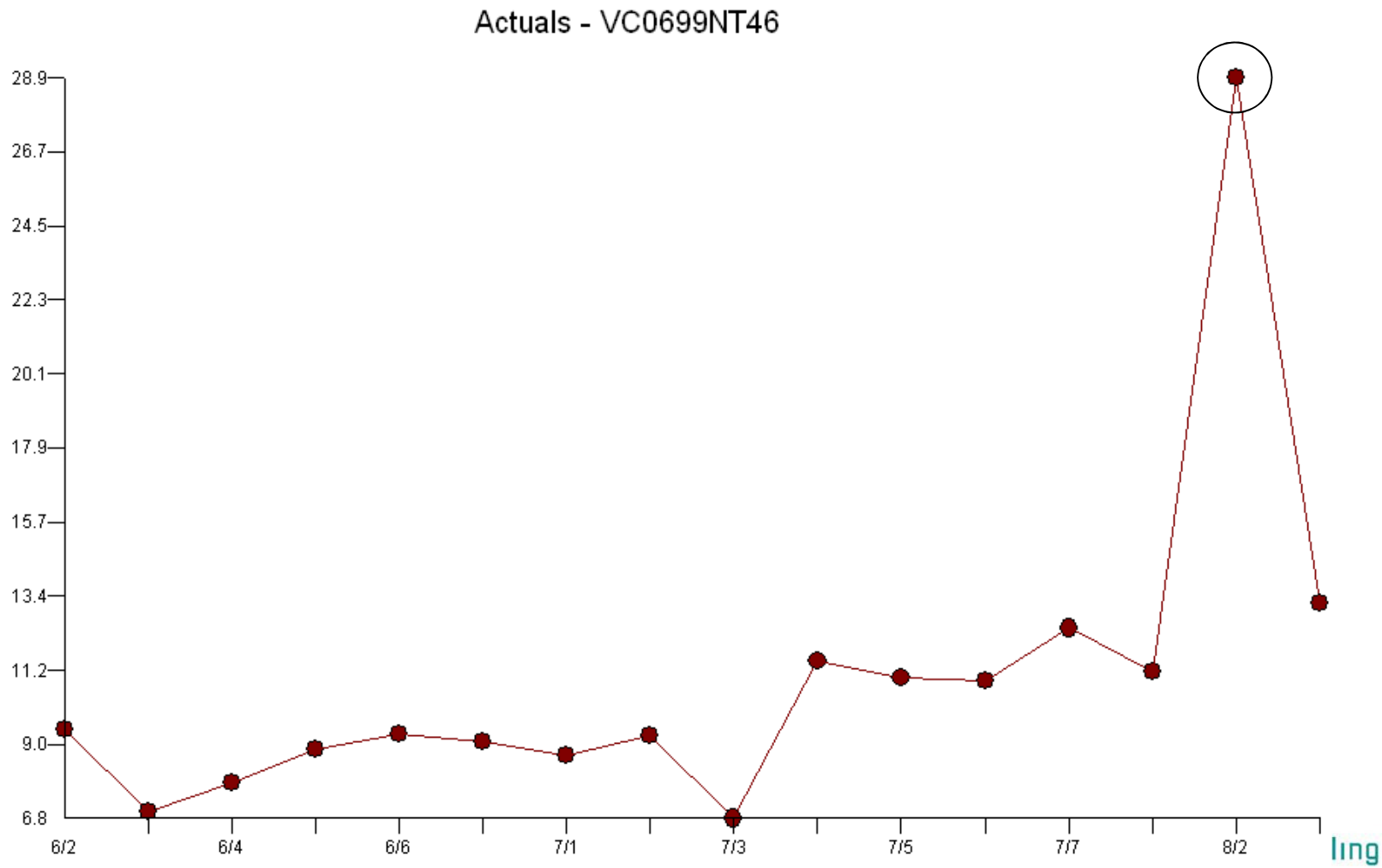
MERCURY MARINE

business modelling ASSOCIATES

AUTOMATIC FORECASTING SYSTEMS

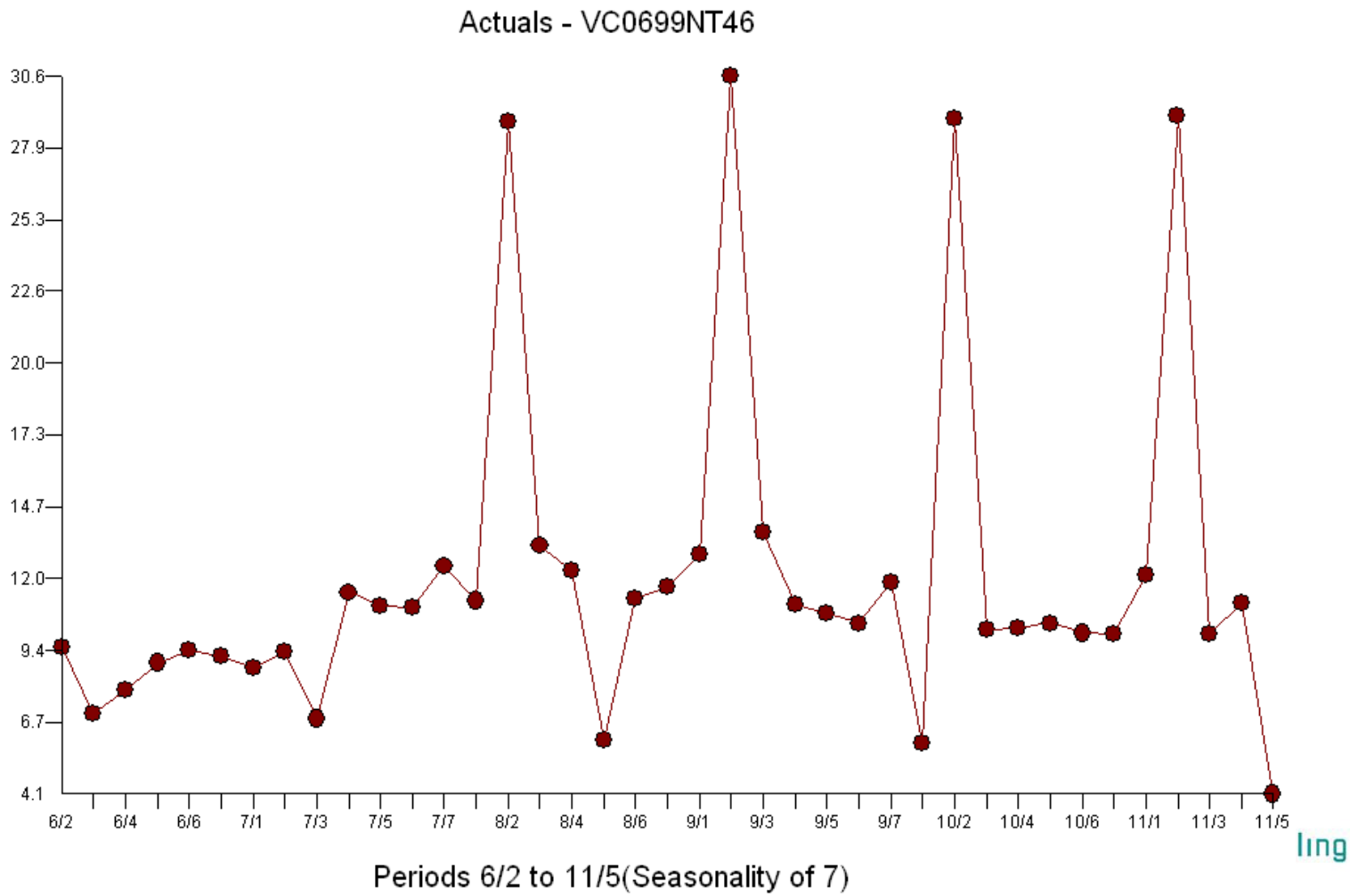
Here is an outlier, right?

- **This value is <u>not</u> an unusual data point**



Seasonal Pulses

- **The outlier is really a series of outliers called a seasonal pulse where the seasonality changed over time**



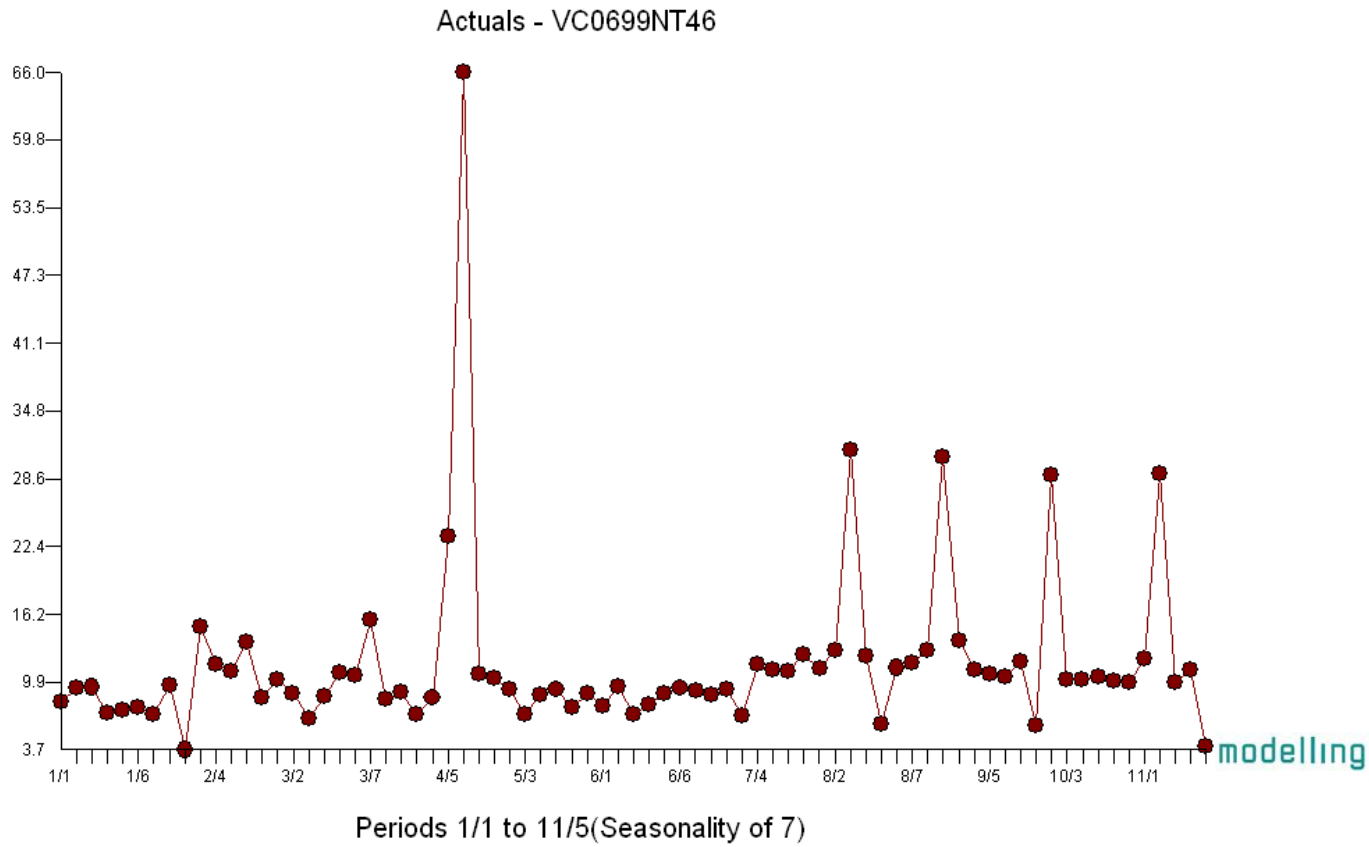
MERCURY MARINE

AUTOMATIC FORECASTING SYSTEMS

What is unusual?

- We see a big outlier in the middle, but what about the pattern near the end?

- Do we remove/fix those also? Do we set them to be an average of the previous and next data points? Or do we identify those as "seasonal pulses" and include them as causal variables in the model so that they can be forecasted correctly?



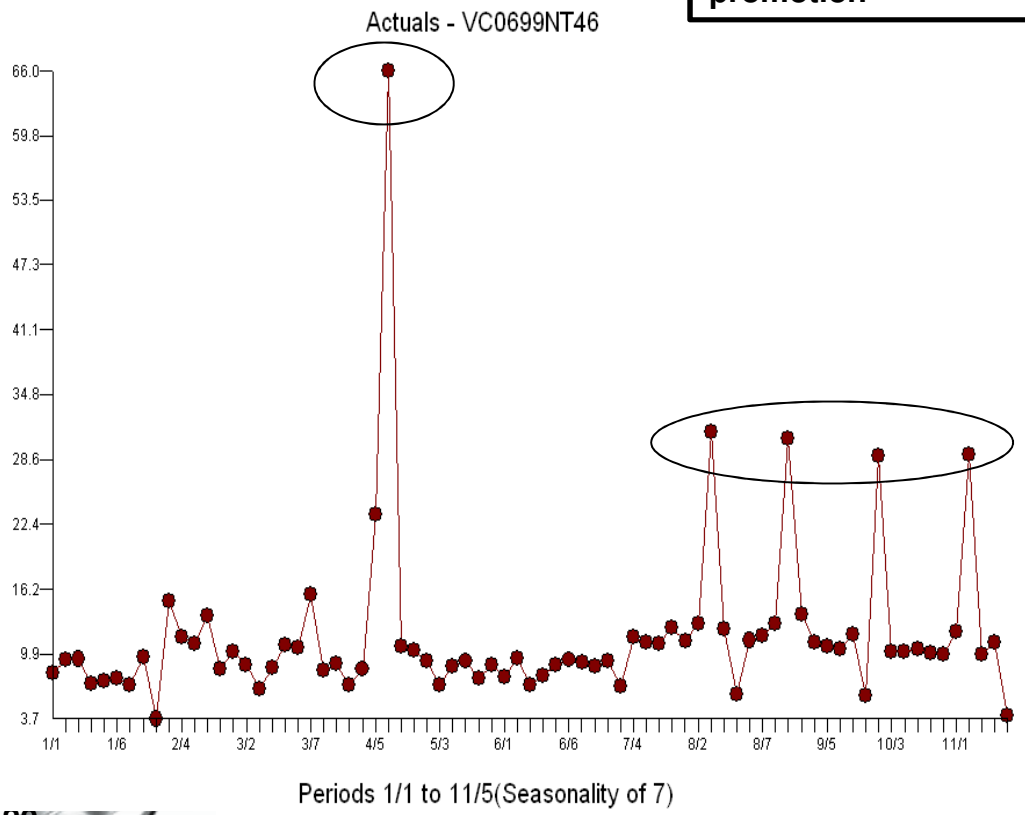
MERCURY MARINE

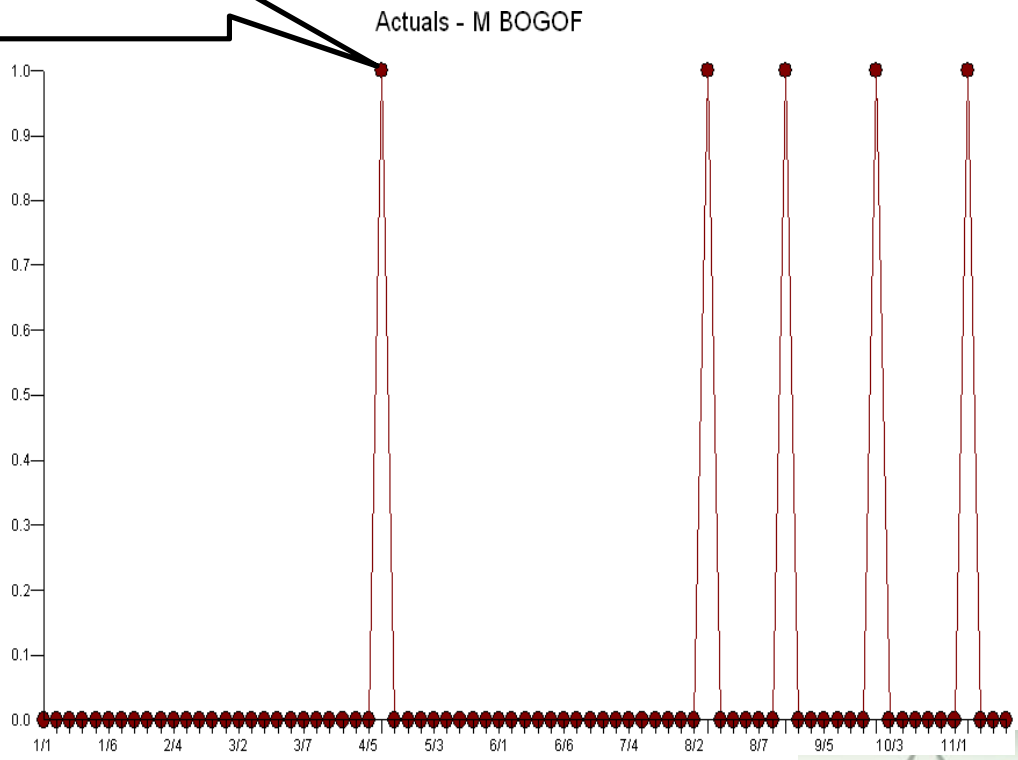
AUTOMATIC FORECASTING SYSTEMS

Was it a Causal Model Issue all along?

- **Oops! We realize that we shouldn't be data cleansing at all. We should be adding causal information to the process. The culprit was that there was a price discount and then a buy one get one free (BOGOF) promotion that caused the change in demand.**

A '1' where there is a promotion and a '0' where there is no promotion





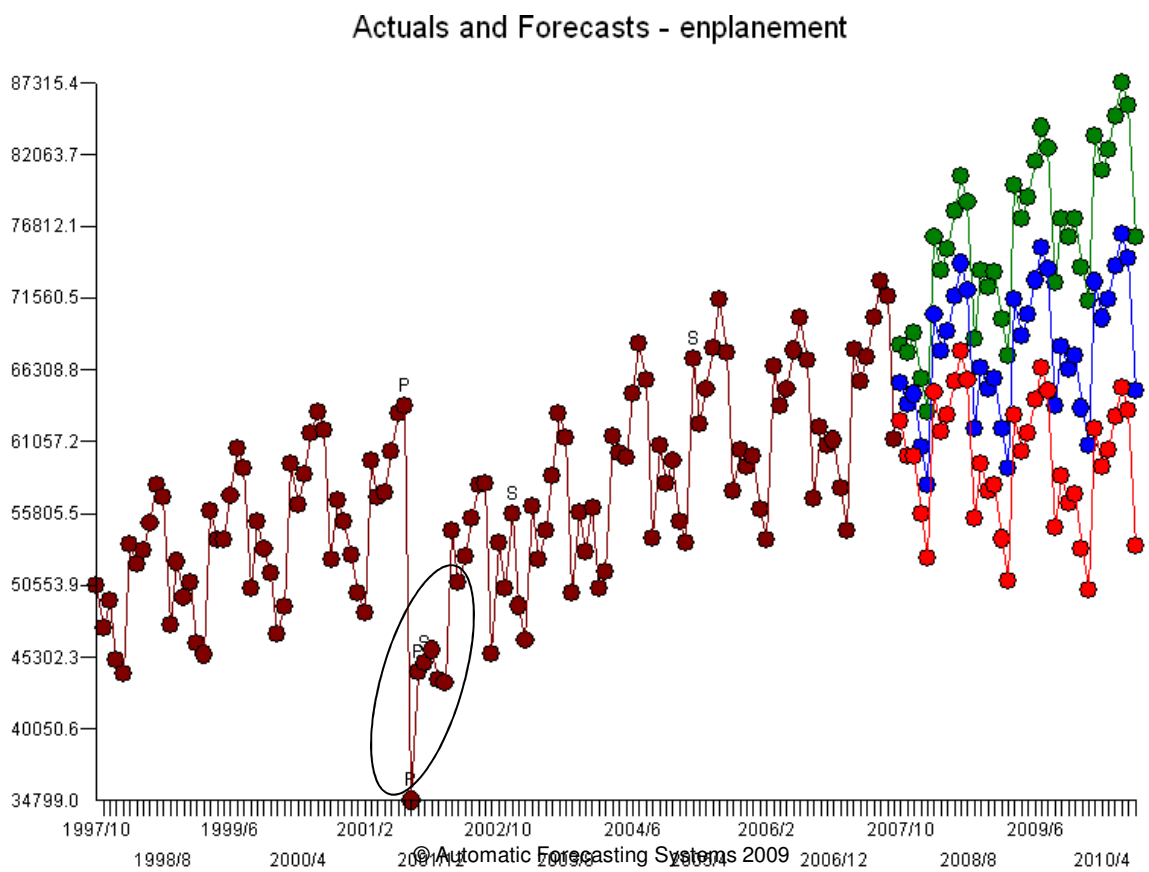
MERCURY MARINE

business modelling ASSOCIATES

AUTOMATIC FORECASTING SYSTEMS

What is unusual?

- This is U.S. enplanement data for 1997-2007. Note the decrease starting 9/11/01. How do you cleanse this dataset? There a few interventions in the fall of 2001 and when corrected for and along with some new seasonal pulses, you are good to go.



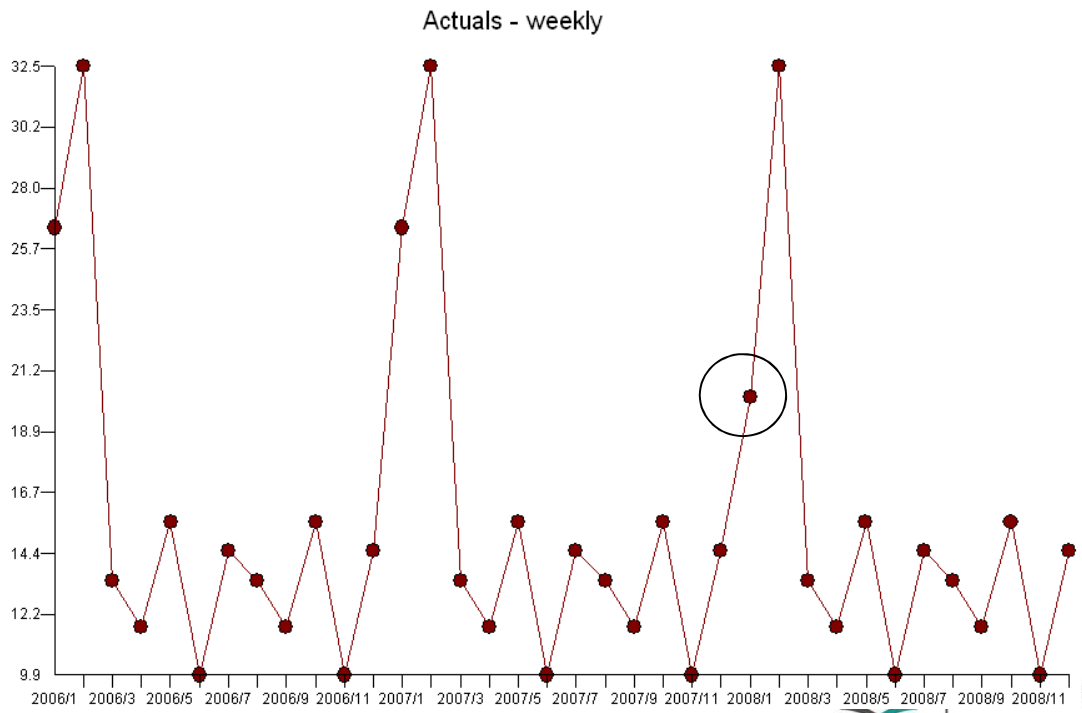
MERCURY MARINE

AUTOMATIC FORECASTING SYSTEMS

What is unusual?

- This is a example where the weekends have high sales. The last Saturday has a low value. Is this an "unusual value"? Yes, but how to identify and account for it. It is an <u>inlier</u> and the remedy is to "tweak" or adjust the observed value to ensure parameter optimization.

- If this value is not accounted for the model parameters and forecast will be affected



MERCURY MARINE

SS modelling ASSOCIATES

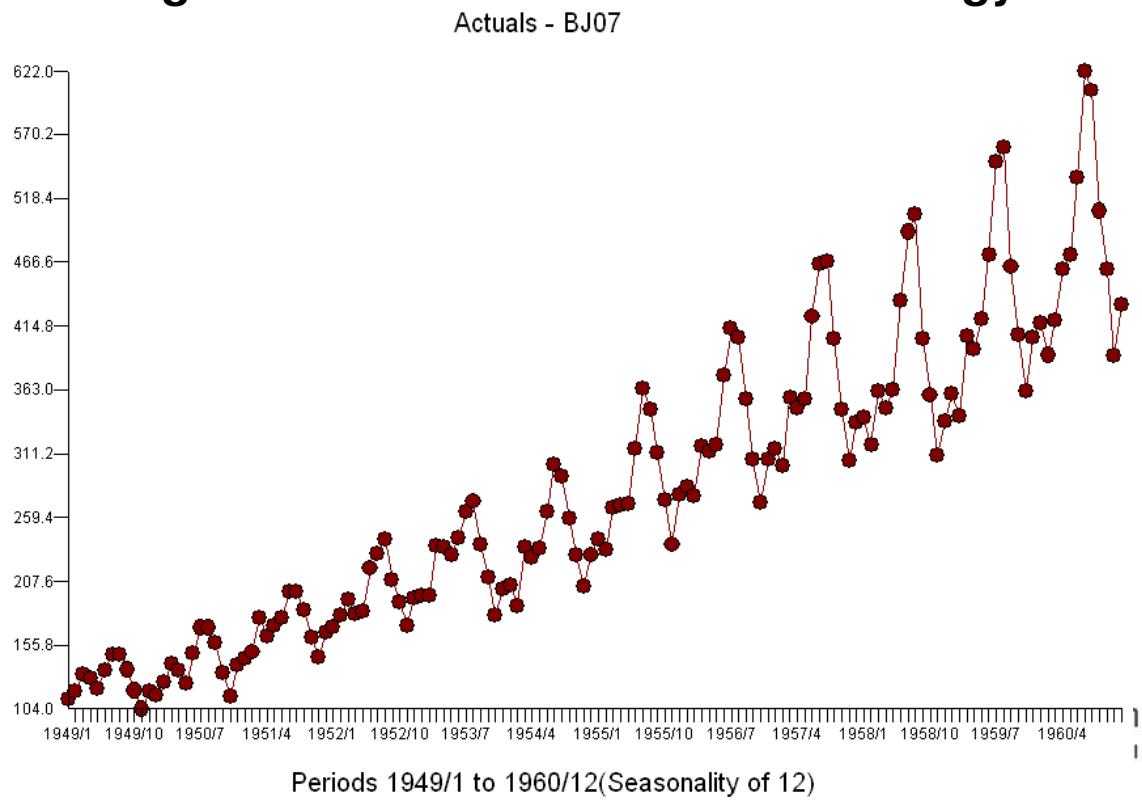
AUTOMATIC FORECASTING SYSTEMS

The “Airline Series”

- One of the most studied time series is the International Airline Passenger’s series(in thousands) for monthly data from 1949 through 1960.

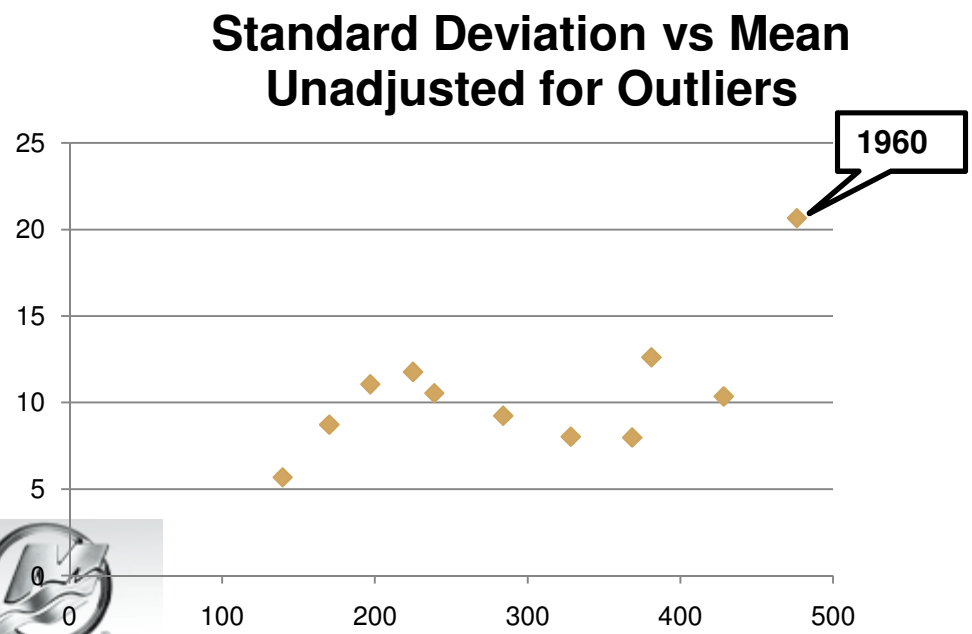
- Box and Jenkins didn’t have the ability to detect outliers and used a log transformation to adjust the data as it seemingly had non-constant variance.

- The forecast was too high and the Box-Jenkins methodology was seen as too complicated.



The "Airline Series"

- The 144 monthly observations were broken into 12 buckets (years) and they calculated the local means (assuming a model) and standard deviations for each bucket (year).
- The conclusion was that the standard deviation was increasing with the mean when it was really outliers in the last year that were skewing the situation by enlarging the standard deviation.



An example of "spurious correlation"

MERCURY MARINE

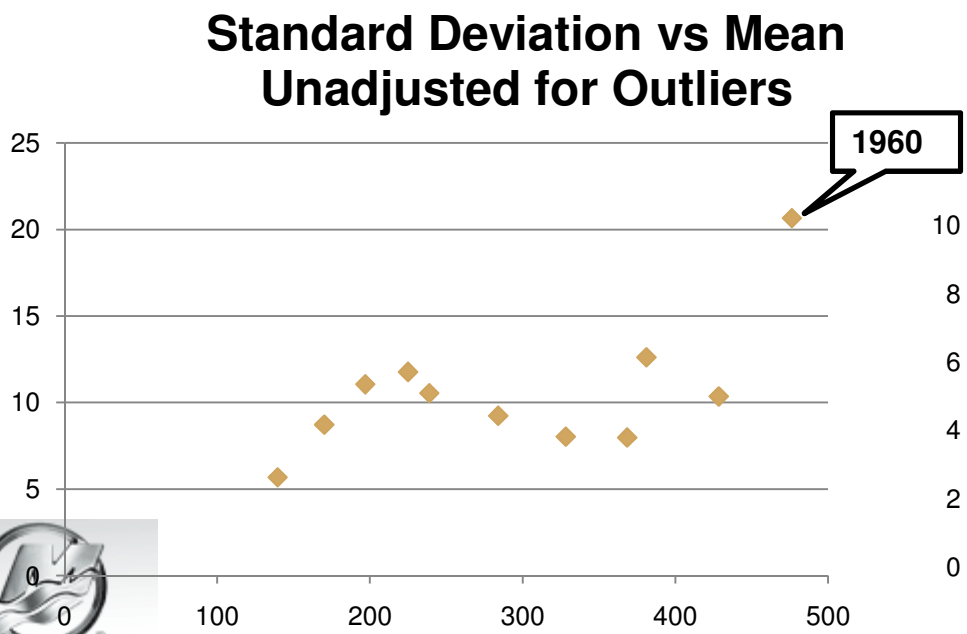
business modelling ASSOCIATES

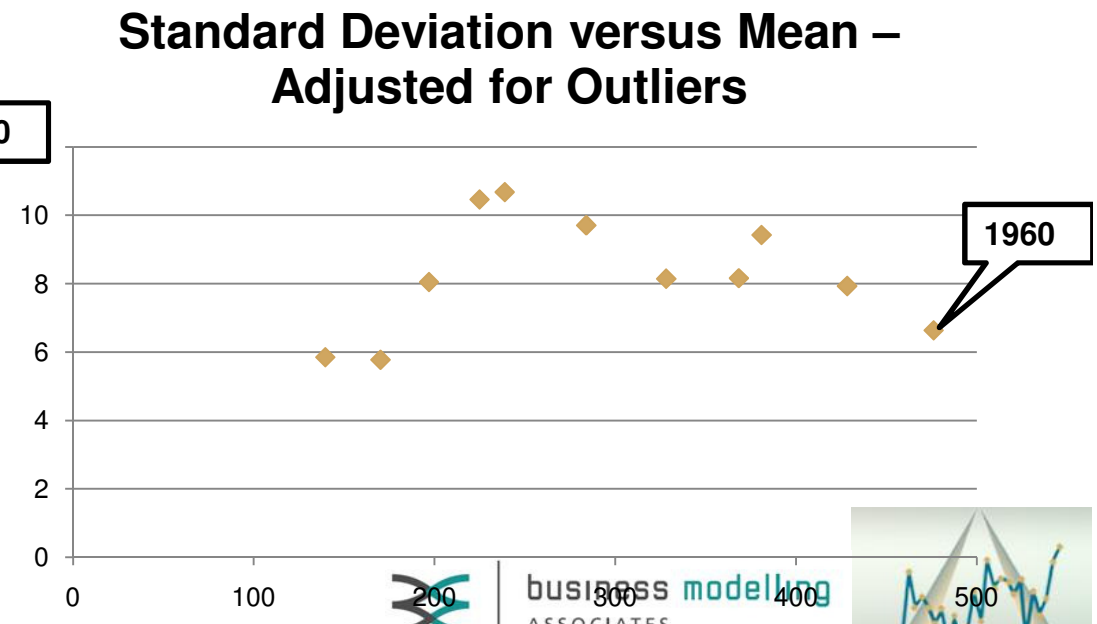
AUTOMATIC FORECASTING SYSTEMS

The “Airline Series”

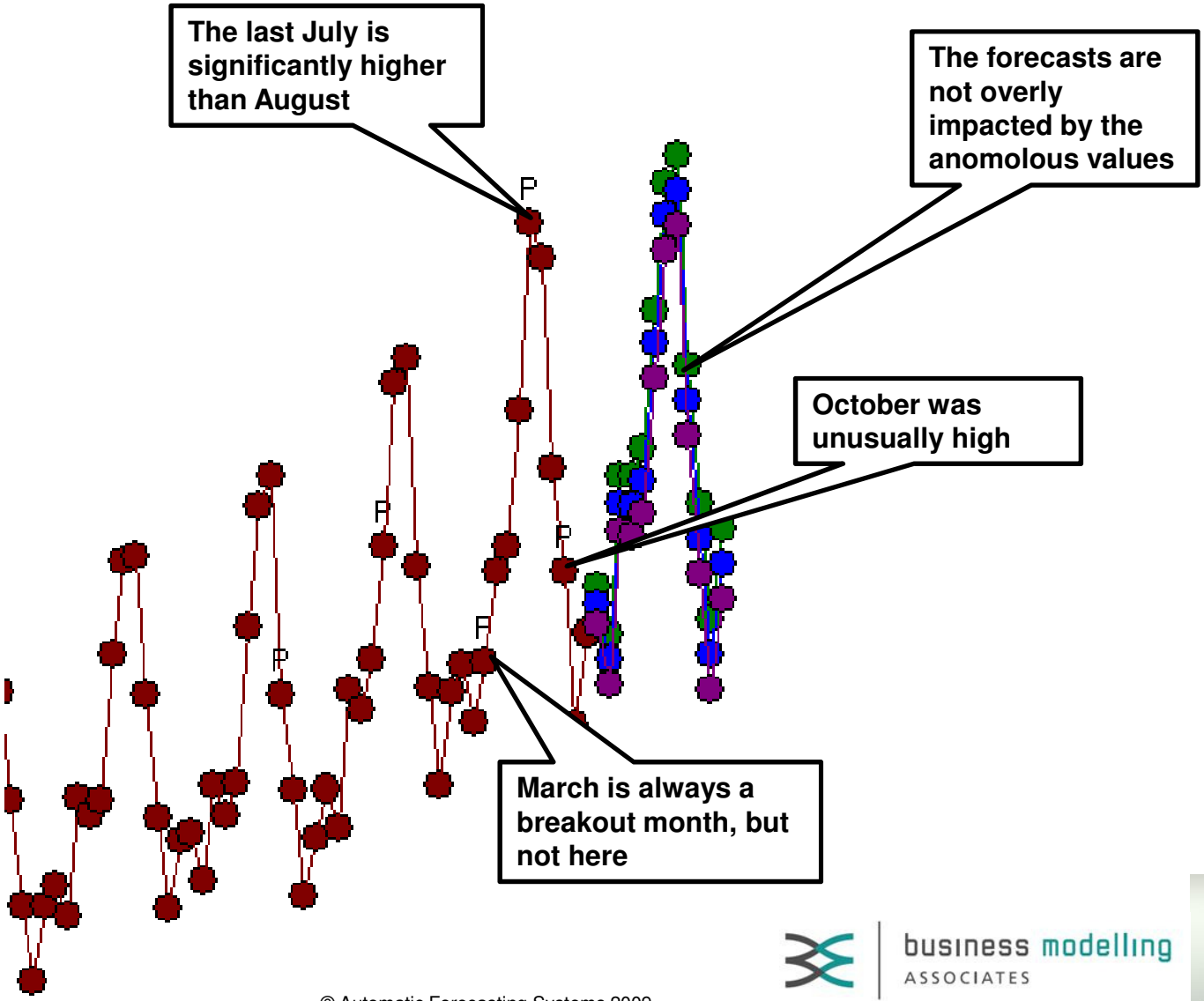
- If we then fit the “airline model”(seasonal differencing and an AR1), identifying and including five outliers (three of them in the last year) we can then use the residuals to calculate the standard deviation for each of the buckets. We then plot the standard deviations against the local means of the observed series and we get another story altogether.

- The conclusion was that the standard deviation was increasing with the mean when it was really outliers in the last year that were skewing the situation.





Does your forecasting software spot the outliers in 1960? If not, ask them why??????



MERCURY MARINE

business modelling ASSOCIATES

AUTOMATIC FORECASTING SYSTEMS

Bad Forecasting Practices still to be found in 2011

- **There are software firms that don't know that instead of taking logs as Box-Jenkins recommended in their 1976 text book, is that log transformations may not be necessary once the data has been cleansed. (Note: Some of today's textbooks also do not know or practice intervention detection !)**

- **SAS**

  - **http://support.sas.com/documentation/cdl/en/etsug/60372/HTML/default/etsug_arima_sect056.htm**

Example 7.2 Seasonal Model for the Airline Series

The airline passenger data, given as Series G in Box and Jenkins (1976), have been used in time series analysis literature as an example of a nonstationary seasonal time series. This example uses PROC ARIMA to fit the airline model, ARIMA(0,1,1)×(0,1,1)$_{12}$, to Box and Jenkins' Series G. The following statements read the data and log-transform the series:

```
title1 'International Airline Passengers';
title2 '(Box and Jenkins Series-G)';
data seriesg;
   input x @@;
   xlog = log( x );
   date = intnx( 'month', '31dec1948'd, _n_ );
   format date monyy.;
datalines;
```

Bad Forecasting Practices still to be found in 2011

- **There are software firms that don't know that instead of taking logs as Box-Jenkins recommended in their 1976 text book, is that log transformations may not be necessary once the data has been cleansed. (Note: Some of today's textbooks also do not know or practice intervention detection !)**

- **Oracle**

  - **http://oracledmt.blogspot.com/2006/03/time-series-forecasting-2-single-step.html**

First we stabilize the variance. This can be done by applying a Box-Cox power transform. This transform has the following form: y(h) = (y^h - 1) / h, if h is not equal to 0 and y(h) = log(y) if h is 0. In general, the LOG transform (h=0) is a good choice for removing increasing variability. Figure 2 shows the transformed series after the LOG transform. The upward trend over time is still visible but the amount of variation in the series is about the same throughout the series.

MERCURY MARINE

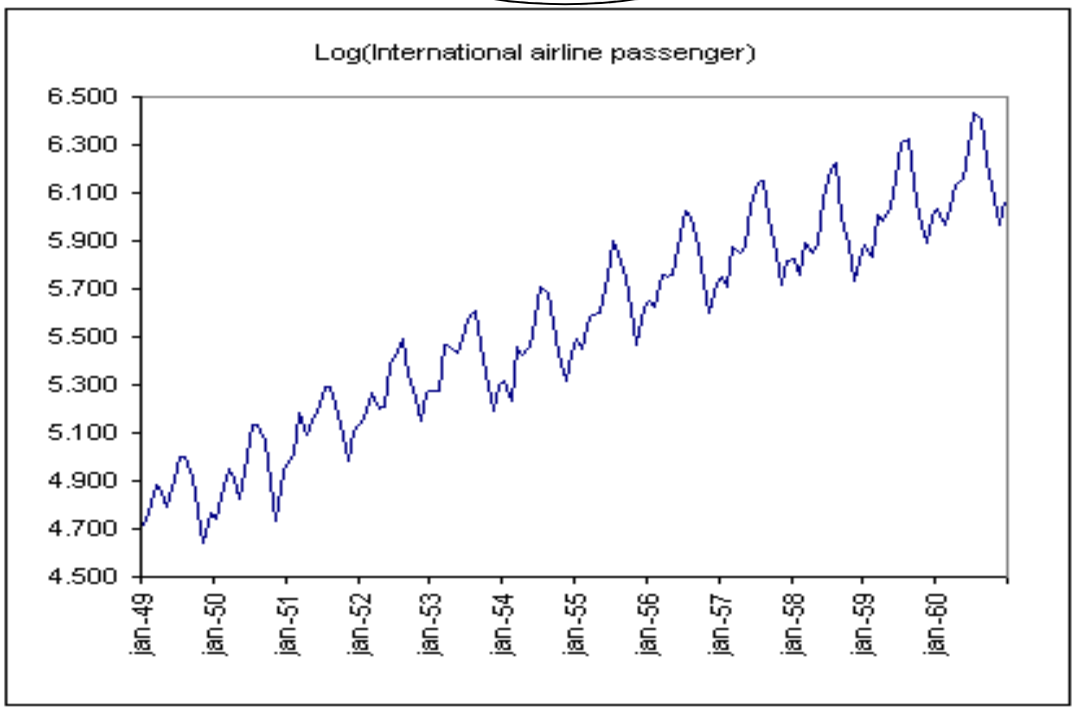
business modelling ASSOCIATES

AUTOMATIC FORECASTING SYSTEMS

Bad Forecasting Practices still to be found in 2011

- There are software firms that don't know that instead of taking logs as Box-Jenkins recommended in their 1976 text book, is that log transformations may not be necessary once the data has been cleansed. (Note: Some of today's textbooks also do not know or practice intervention detection!)
- XLSTAT
  - http://www.xlstat.com/en/support/tutorials/arima.htm

We notice on the chart, that there is a global upward trend, that every year a similar cycle starts, and that the variability within a year seems to increase over time. Before we fit the ARIMA model, we need to stabilize the variability. To do that, we transform the series using a log transformation. We can see on the chart below that the variability is reduced.



Bad Forecasting Practices still to be found in 2011

- **There are software firms that don't know that instead of taking logs as Box-Jenkins recommended in their 1976 text book, is that log transformations may not be necessary once the data has been cleansed. (Note: Some of today's textbooks also do not know or practice intervention detection !)**

- **Visual Numerics (IMSL- International Math and Statistics Library)**

  - **http://www.vni.com/products/imsl/documentation/fort06/stat/NetHelp/default.htm?turl=bctr.htm**

Example 1

Consider the Airline Data (Box and Jenkins 1976, page 531) consisting of the monthly total number of international airline passengers from January 1949 through December 1960. Routine BCTR is used to compute a forward Box-Cox transformation of the first 12 observations. In the transformation SHIFT and POWER are each set to zero, which corresponds to taking natural logarithms of the data.

MERCURY MARINE

business modelling ASSOCIATES

AUTOMATIC FORECASTING SYSTEMS

Bad Forecasting Practices still to be found in 2011

- **There are software firms that don't know that instead of taking logs as Box-Jenkins recommended in their 1976 text book, is that log transformations may not be necessary once the data has been cleansed. (Note: Some of today's textbooks also do not know or practice intervention detection!)**

- **MATLAB**

  - **http://www.mathworks.com/products/statistics/demos.html?file=/products/demos/shipping/stats/stattsdemo.html**

Examine Trend and Seasonality

This series seems to have a strong seasonal component, with a trend that may be linear or quadratic. Furthermore, the magnitude of the seasonal variation increases as the general level increases. Perhaps a log transformation would make the seasonal variation be more constant. First we'll change the axis scale.

```
set(gca,'YScale','log');
```


MERCURY MARINE

business modelling ASSOCIATES

AUTOMATIC FORECASTING SYSTEMS

Bad Forecasting Practices still to be found in 2011

- **There are software firms that don't know that instead of taking logs as Box-Jenkins recommended in their 1976 text book, is that log transformations may not be necessary once the data has been cleansed. (Note: Some of today's textbooks also do not know or practice intervention detection!)**

- **Mathematica**

  - **http://media.wolfram.com/documents/TimeSeriesDocumentation.pdf**

We see that the variance, riding on the trend, is also changing with time. We need to transform the series into a constant variance before modeling it further. To stabilize the variance, a nonlinear transformation such as a logarithmic or square-root transformation is often performed. In this example, we try a natural logarithmic transformation, $y_t = \ln x_t$.

This is the time plot of the airline data after the logarithmic transformation. Note that `Log[aldata]` gives the logarithm of each of the entries of `aldata`.

```
In[40]:= ListLinePlot[Log[aldata], AxesLabel -> {"t", "ln(xt)"}]
```

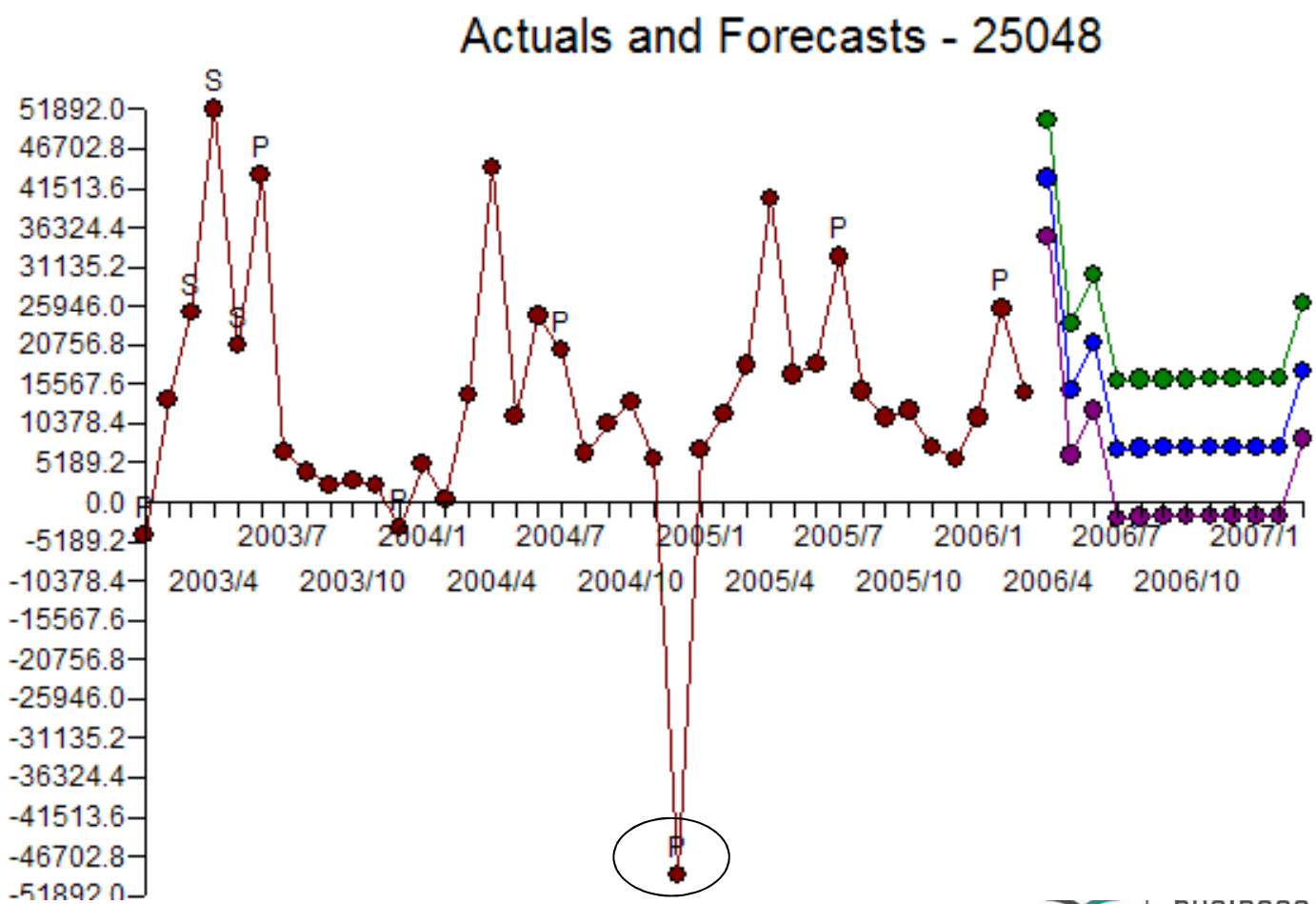

MERCURY MARINE

business modelling ASSOCIATES

AUTOMATIC FORECASTING SYSTEMS

Outlier Detection – Pulse

- **Pulse – Fire in the warehouse in April (0,0,0,0,0,0,0,0,1,0,0,0,0,0)**



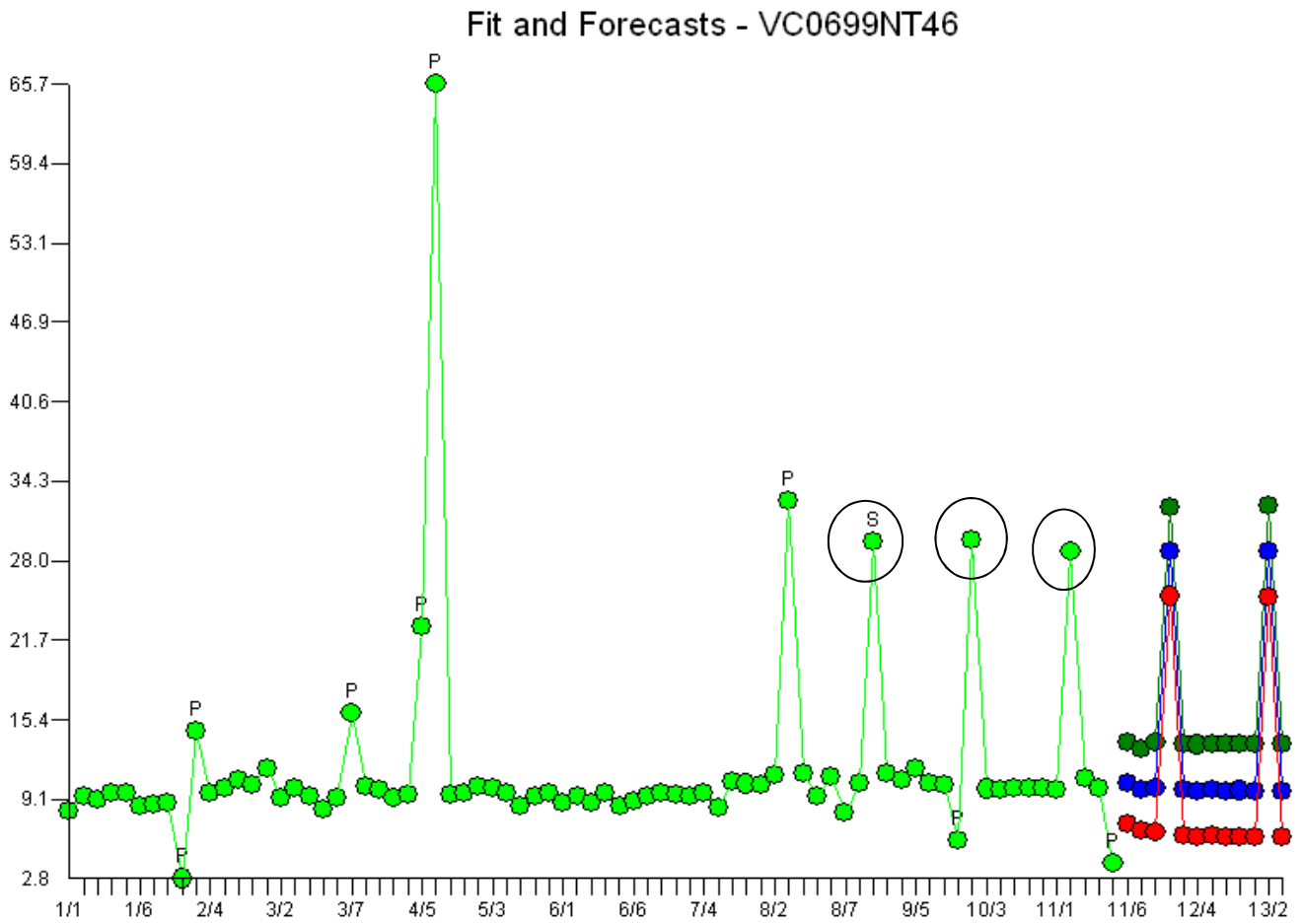
MERCURY MARINE

business modelling ASSOCIATES

AUTOMATIC FORECASTING SYSTEMS

Outlier Detection – Seasonal Pulse

- **Seasonal Pulse – February emerges later during the year (0,1,0,0,0,0,0,0,0,0,0,0,0,1)**



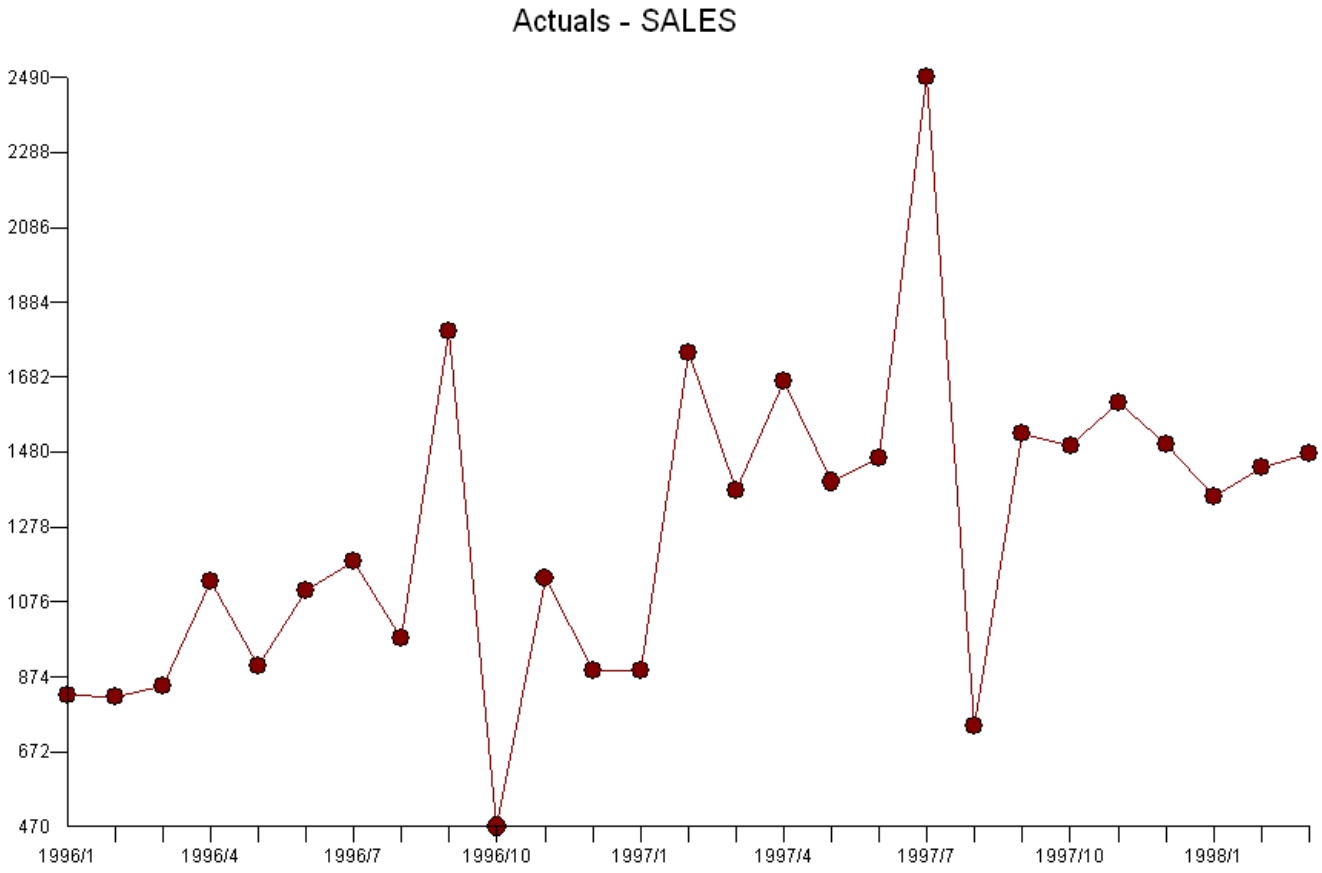
MERCURY MARINE

modelling ASSOCIATES

AUTOMATIC FORECASTING SYSTEMS

Tough Series to Model

- **From a visual it looks like a seasonal model that is increasing, right?**



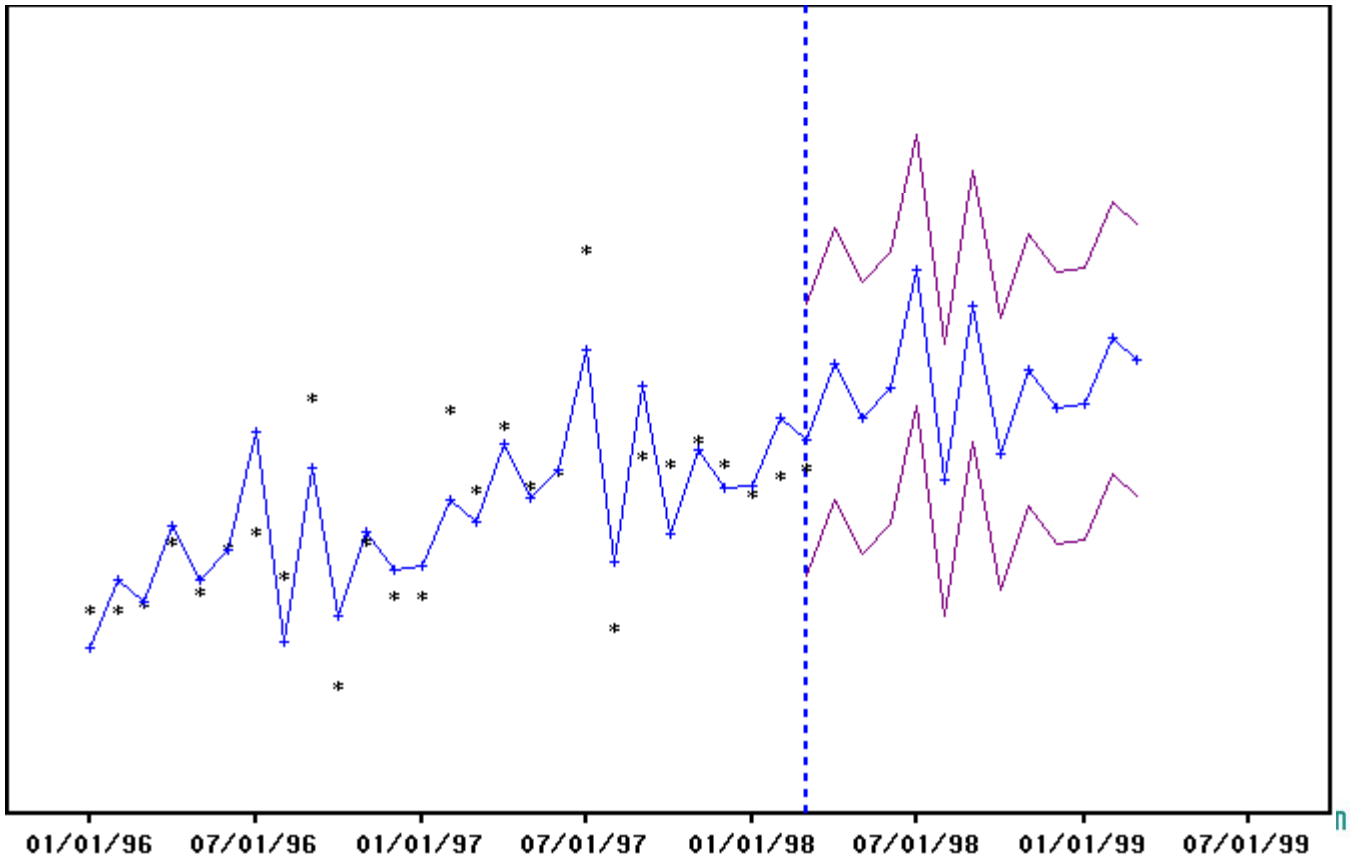
MERCURY MARINE

business modelling ASSOCIATES

AUTOMATIC FORECASTING SYSTEMS

What Would PROC Do?

- Drive a Holt-Winters Model through its heart and predict an upwards trend?
- Be fooled that there is seasonality even though the blips are 13 periods apart not 12?

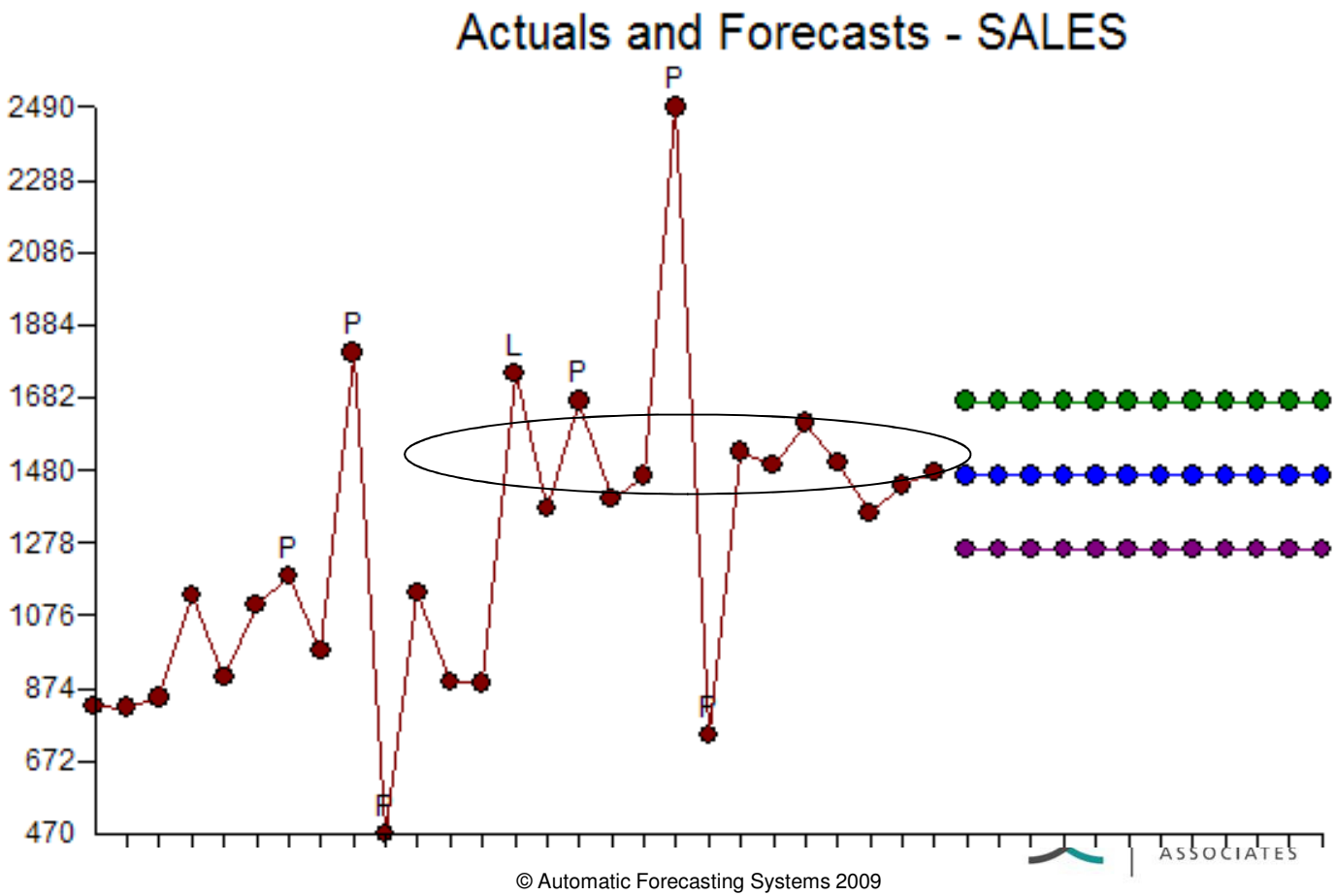


MERCURY MARINE

AUTOMATIC FORECASTING SYSTEMS

Outlier Detection – Level Shift

- **Level Shift –Competitor drops out of the market and an 'one-time' increase in market share gain (0,0,0,0,1,1,1,1,1,1,1,1)**

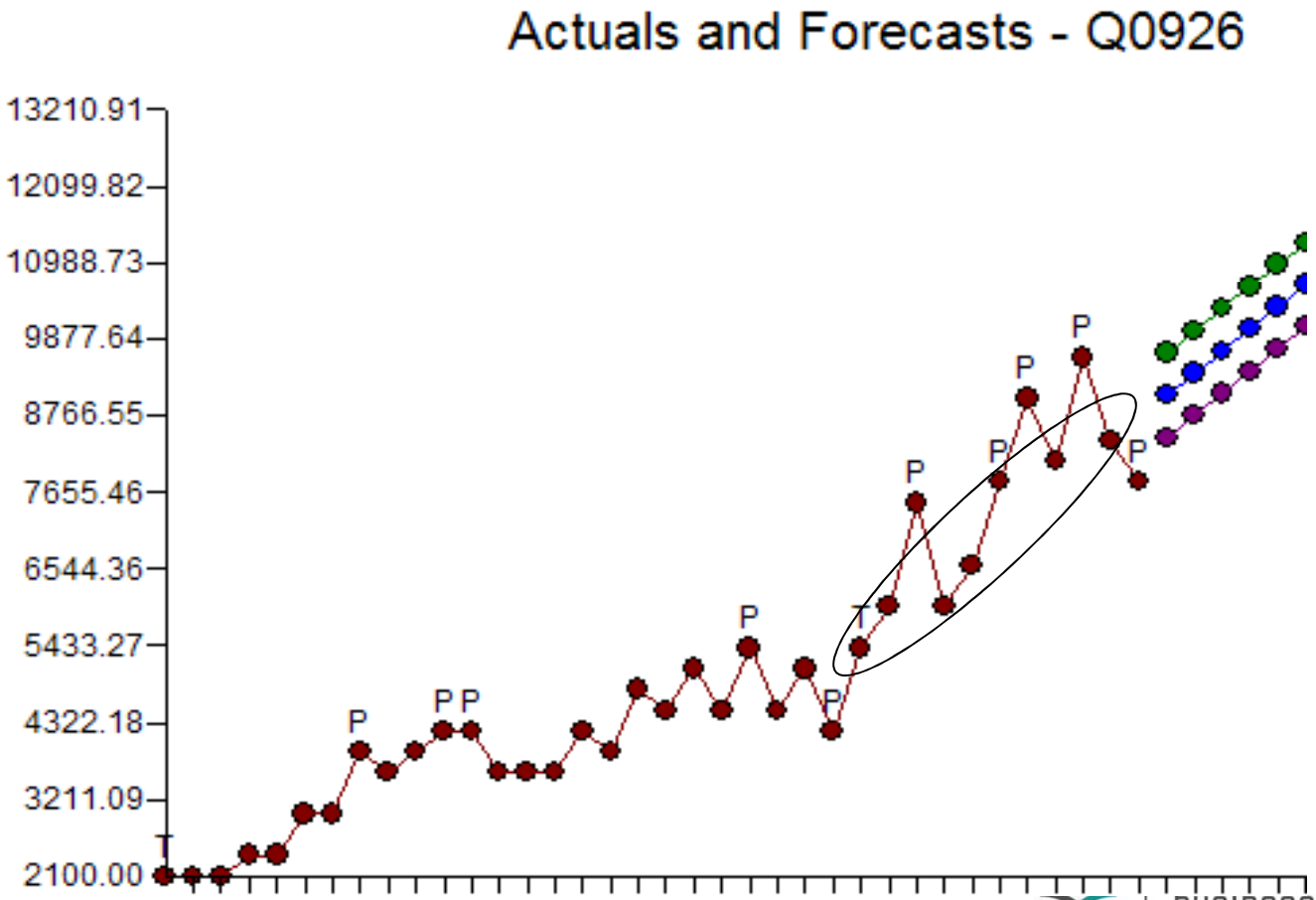


MERCURY MARINE

AUTOMATIC FORECASTING SYSTEMS

Outlier Detection – Local Time Trend

- **Local Time Trend – A new trend up or down very different from the past (0,0,0,0,1,2,3,4,5,6,7,8,9,etc.)**



MERCURY MARINE

business modelling ASSOCIATES

AUTOMATIC FORECASTING SYSTEMS

Outlier Detection – What should you do about it?

- User Provides knowledge - before the modelling process begins – If there is some domain knowledge that there was an event in the past then this information should be included in the model as a possible input variable so that the observed value is not "adjusted". In this case an actual variable now has a coefficient and can explain the impact so that it's effect can be anticipated in the future if the candidate variable is operational or in effect.

- Action - You don't want to believe a pulse and you should adjust the pulse to "where it should have been" thus providing a robust estimation of the model parameters.

- No Action - If you do not adjust for outliers then the coefficients in the model will be skewed creating a false image of the systematic behavior. The forecast may be higher or lower than anticipated. The causal relationship may have an incorrect snapshot of the relationship between price and sales for example.

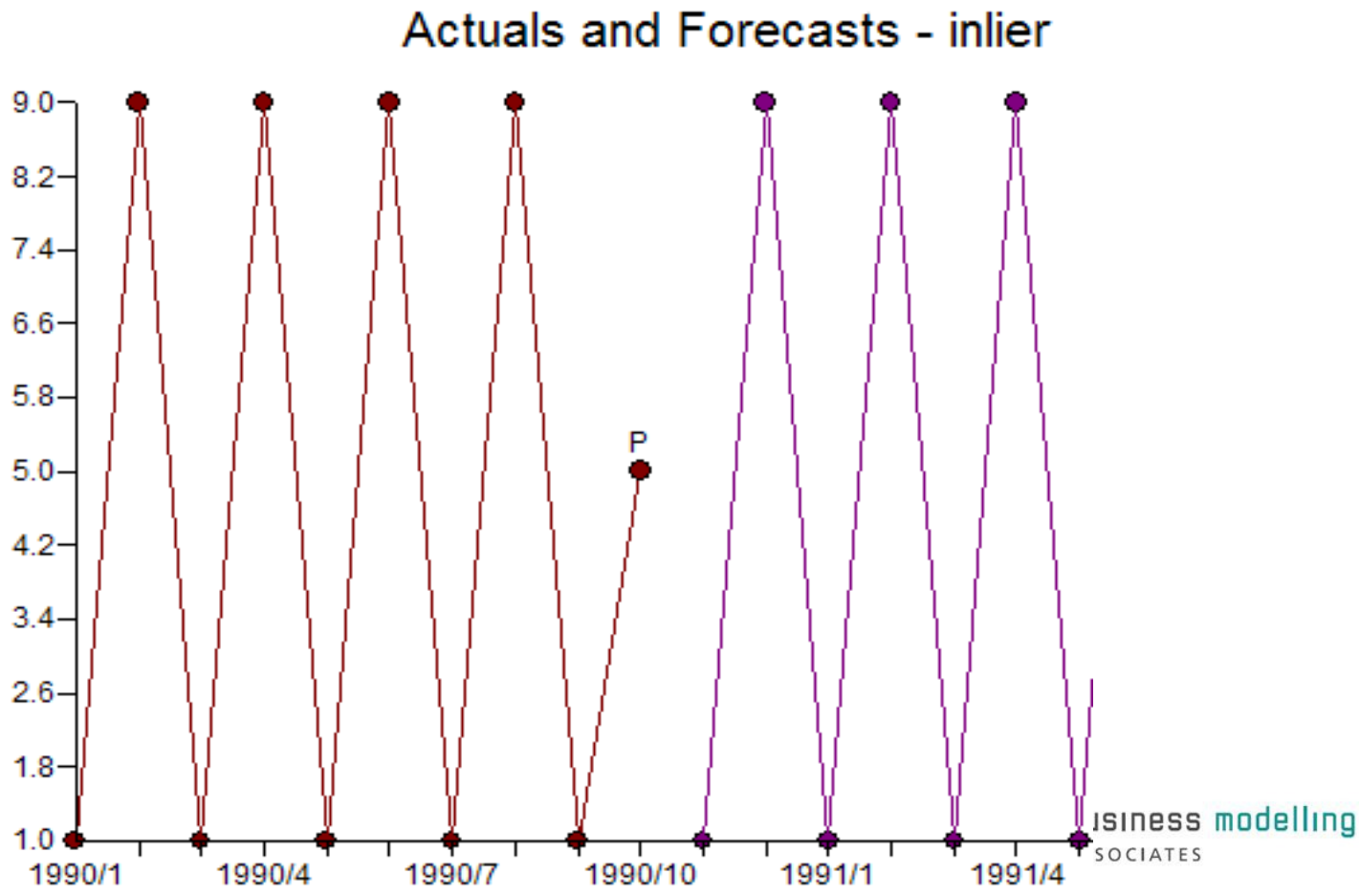
MERCURY MARINE

business modelling ASSOCIATES

AUTOMATIC FORECASTING SYSTEMS

Outlier Detection

- **Keeping what is unusual within the standard deviation bound:**
  - **If a pulse occurs that is an "inlier" and not outside the standard deviation bounds then it will not be identified as you need a model first to identify this situation (e.g. 1,9,1,9,1,9,5)**



Outlier Detection - Out of Model Approach

- Calculate Residuals (Actual - Mean)
- Calculate the standard deviation around the mean
- Specify the # of standard deviations around the mean that will be considered an outlier (e.g. 3)
- Identify those observations outside the standard deviation
- Replace the unusual observations with the mean or interpolation
- Specify the number of iterations to go through the outlier removal process

Outlier Detection - Within Model Approach

- Identify a possible model for the data.
- Identify the outlier in the presence of the model effect.
- A statistically valid test is performed on the unusual value which uses a "standard deviation of the error" that excludes the impact of the outlier and therefore is more robust
- Consider identifying outliers first and then the model form.
- Evaluate the alternatives of model first then outlier detection. Compare results and determine the optimal strategy to follow for each dataset.

MERCURY MARINE

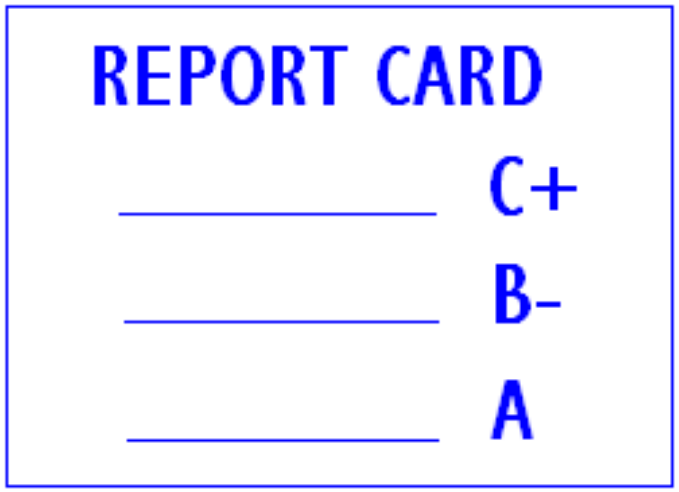
business modelling ASSOCIATES

AUTOMATIC FORECASTING SYSTEMS

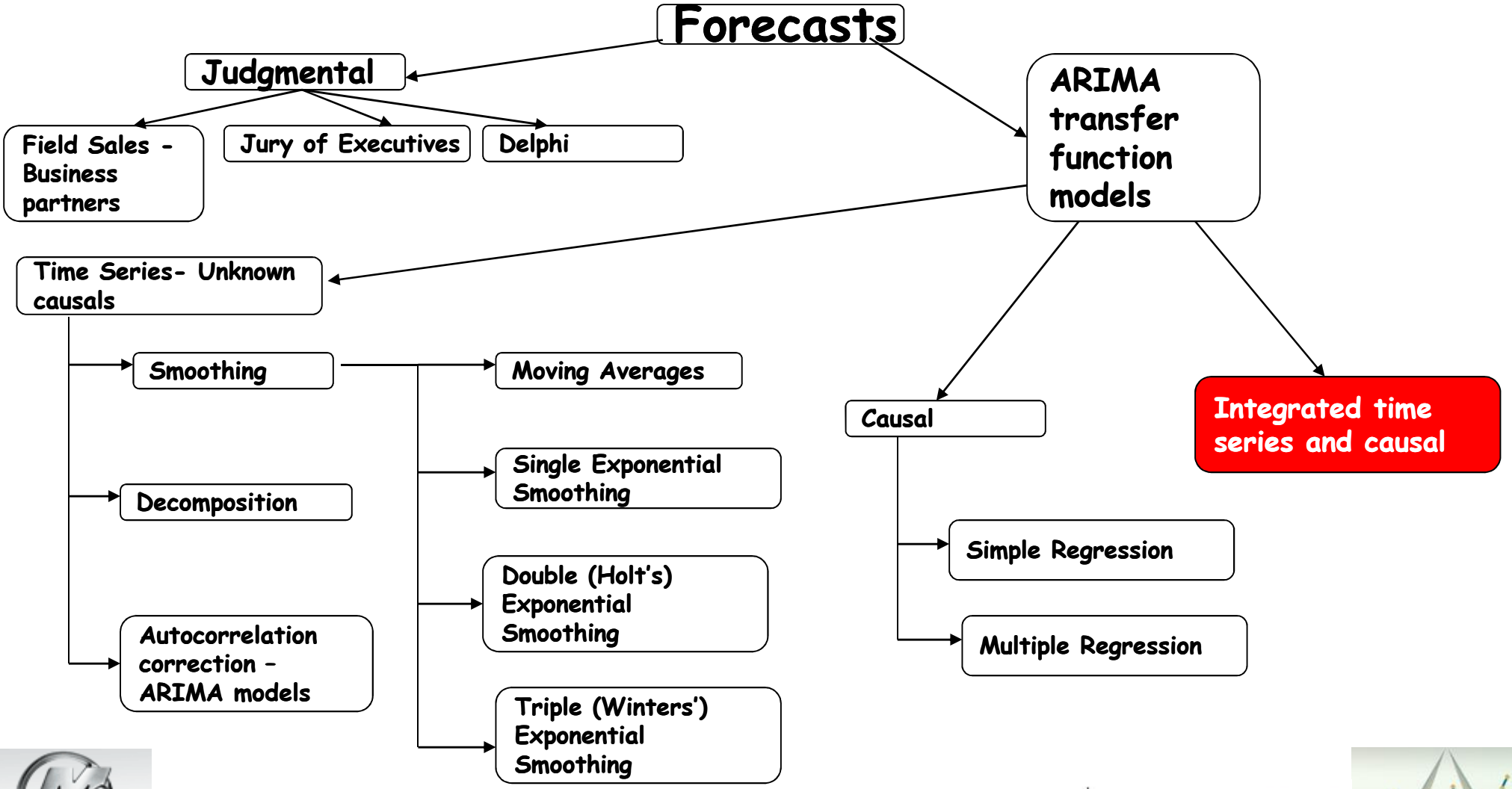
AUTOMATIC FORECASTING SYSTEMS

The dawn of a new age in forecasting!

A Critique of Automatic Forecasting Software



Forecasting Methods Family Tree



MERCURY MARINE

business modelling ASSOCIATES

AUTOMATIC FORECASTING SYSTEMS

History of Modelling

- 1795 - Legendre – Regression - Developed for cross-sectional data and later abused using with time series data – You can swap first observation with last observation and get the same answer
- 1920 - Seasonal Decomposition
- 1927 - Slutsky-Yule – Identified that applying a moving average to a random process may in itself create a pattern when no existed previously
- 1944 - Brown – Exponential Smoothing
- 1957 - Holt – Holt Method
- 1960 - Holt/Winters – Holt/Winters
- 1960 - Chow test – Parameter Changes
- 1965 - Almon – Polynomial Distributed Lag
- 1967 - Hiskin/Young, Musgrave – Seasonal Adjustment Census Bureau
- 1970 - Box & Jenkins – Time Series Analysis Textbook –
  - Introduced a Generic model form which *all* models are a subset
  - Introduced a "Data based" approach of building a unique model and coefficients for each data set
- 1976 - Box and Tiao – Interventions
- 1988 - Chang, Tiao and Chen – Innovational Outliers
- 1988 - Tsay – Level Shifts, Variance Change

"Model based" assuming the relationship within the data is a certain weighting scheme and the number of periods to weight.

MERCURY MARINE

business modelling ASSOCIATES

AUTOMATIC FORECASTING SYSTEMS

Progress in a Real World View of History of Modeling

- Let's use the last 100 days to predict tomorrow's rainfall using an average
- Let's only use the last 12 days and weight the more recent data more and the older data less
- Let's use only the last 7 days using a weighting scheme and provide a bump up on Friday's as it rains more on Friday's
- Let's use an "additive" method to adjust by way off addition or subtraction for the forecast fluctuations
- Let's use a "multiplicative" method to adjust by way off %'s up and down for the forecast fluctuations
- By using the Box-Jenkins approach of calculating lags of the history using regression to identify the length of time to use and the weighting instead of assuming
- By using Intervention Detection, you can add unspecified causal variables to the model that adjust for outliers, level shifts, local time trends and seasonal pulses that if not accounted for will distort the coefficients in the model and thereby the forecasts

MERCURY MARINE

business modelling ASSOCIATES

AUTOMATIC FORECASTING SYSTEMS

Box-Jenkins Modeling Steps

Identification - Calculate statistics on the data to suggest a model form (length and weight) using the ACF & PACF

Estimation - Taking suggested model form and estimating the optimal coefficients

Diagnostic Checking - Making sure that the residuals are constant mean/variance, random and no autocorrelation

Forecasting - Take the estimated model and generate X period out forecasts

MERCURY MARINE

business modelling ASSOCIATES

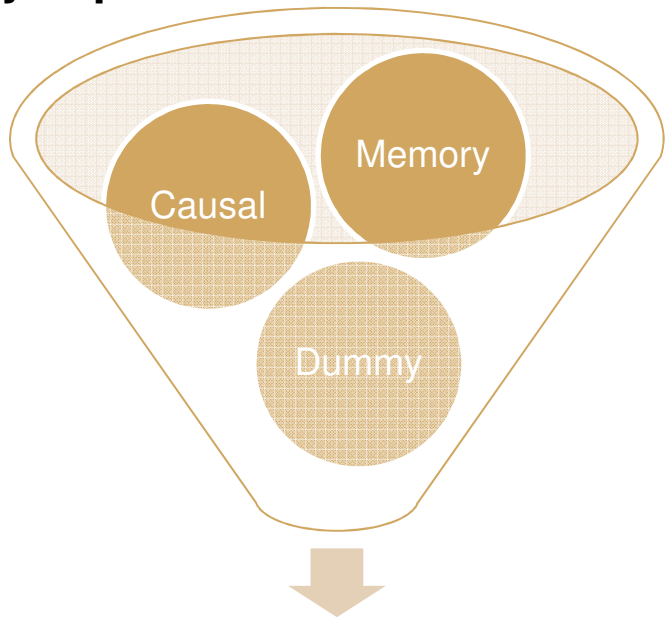
AUTOMATIC FORECASTING SYSTEMS

Regression Modelling

- **Causal variables - Variables like Price might have a lead or lag relationship and that exact period may be difficult to identify. Assuming it is not going to help you.**

- **Dummy variables – Outliers need to be provided and/or identified and adjusted for by the system.**

- **Memory – There is a period to period relationship that exists in the data. The historical data implicitly captures the effect of omitted causal variables.**



MERCURY MARINE

business modelling ASSOCIATES

AUTOMATIC FORECASTING SYSTEMS

Regression Modelling

- Most Software packages allow a user to provide causal variables like price and promotion and events (e.g. holidays, outliers), but <u>assuming</u> that the relationship between the causals and the output series are current with no lead/lag

- "Skipping Identification" and going right into Estimation means that the lag or lead relationship between the causals and the output series has not been identified
  - Some will assume a theoretical 3 month lag relationship, but is it?
  - Some will attempt to plot lags of causals vs. the output variable to "see" the lag or lead relationship
  - Some with more knowledge will use methods to use a statistic called the "cross-correlation function" to identify the relationship
  - Some may review the residuals from the model and then add in an ARIMA of lag 1 to correct for period to period relationship of the output variable (e.g. Hildreth-Liu or Cochrane-Orcutt) instead of identifying the model

- Trying to Identify the relationship takes more computing time
  - Is there a lead or lag relationship?
  - Are there are outliers that need to be incorporated?

MERCURY MARINE

business modelling ASSOCIATES

AUTOMATIC FORECASTING SYSTEMS

Automatic Modelling

- User might be asked(forced?) to order the data into groupings before any modelling ever occurs so the modeling process doesn't get "fooled":
  - Is the data seasonal?
  - Is there intermittent demand?
  - Do you have any events (promotions, interventions)?

- System tries various quadratic equations to get the best fit, but lacks any ability to forecast.

- User specifies the model and the system estimates the optimal coefficients.

- User specifies which of the different criteria(i.e. smallest AIC, BIC, SIC, RMSE etc.) to be used to determine the "goodness of fit" from a pre-specified list of models using a withheld number of observations yielding strange results like seasonal models when there is no seasonality in the data at all.

- Intelligent software has heuristics to determine model form, variables that are significant and suggests and includes interventions into the model.

MERCURY MARINE

business modelling ASSOCIATES

AUTOMATIC FORECASTING SYSTEMS

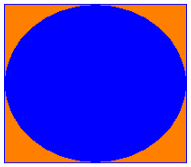
MODELLING VS FITTING

MERCURY MARINE

business modelling ASSOCIATES

AUTOMATIC FORECASTING SYSTEMS

Model Fitting

- Fast and easy to do, but not likely to match the "fingerprint to the killer"
- Fitting a round peg into a square hole 
- Can a "pick-best" approach work in an infinite sample space?
  - The "fitters" will take a list of ~10, 25, 50 models and try to find the model that best "fits" the data. The process will then tell you that it is "optimizing" the parameters, but it just can't be as you assumed a model to begin the process
- Sometimes the "fitters" model get fooled by only partially describing the data
  - Seasonal dummy model used with data that only has <u>some</u> seasonal months
  - A level shift is thought to be a series with an upward trend

MERCURY MARINE

business modelling ASSOCIATES

AUTOMATIC FORECASTING SYSTEMS

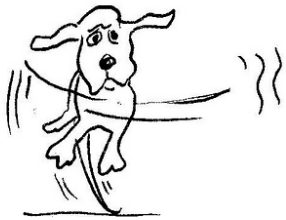
"Gaming the System" by Way of Withholding Data

- We could find which model and which number of observations that are withheld would result in the best fit if we really wanted to, but should you?
- If you have ~50 models and fit them using 1 period withheld
- If you have ~50 models and fit them using 2 periods withheld
- And so on until you have done up until 12 periods withheld
- You would have the combination of one type of model and a certain number to be withheld that would be the "winner" of the best outcome (e.g. model 17 and 4 withheld)
- The reality is that while this exercise in futility would certainly yield the smallest fitting statistic, but it really is not capturing the pattern in the data and is only an exercise "mathematical manipulation"

More on Withholding Observations

- Some will allow users to specify the number of observations to withhold to allow the model "train" so that it optimizes the model for this withhold set of data

- It is the case of the tail wagging the dog



- How do you know how many observations to withhold and what happens if I change the withhold from 6 to 7, will my model and forecast change? You betcha!

- This approach builds a model on data and then changes the coefficients based on the most recent data.

- What if there are outliers not captured by the process in the withhold data? It will skew the model and forecast

- Are the older data worthless? They are rendered so as the withheld data is used to determine the "best" coefficients

MERCURY MARINE

business modelling ASSOCIATES

AUTOMATIC FORECASTING SYSTEMS

Customized Modeling

- Much slower and requires complicated schemes to sift through the patterns in the data to build a customized model for the each data set.

- Did it get the model 100% right? Probably not, but then again the fitter didn't even try and took a passive rather than active approach. It's like passing to the other family in the TV GAME SHOW "Family Feud" instead of trying to answer the question yourself

- It's like getting a custom made suit that fits your dimensions, oops bad graphic!

MERCURY MARINE

business modelling ASSOCIATES

AUTOMATIC FORECASTING SYSTEMS

Caveat Emptor

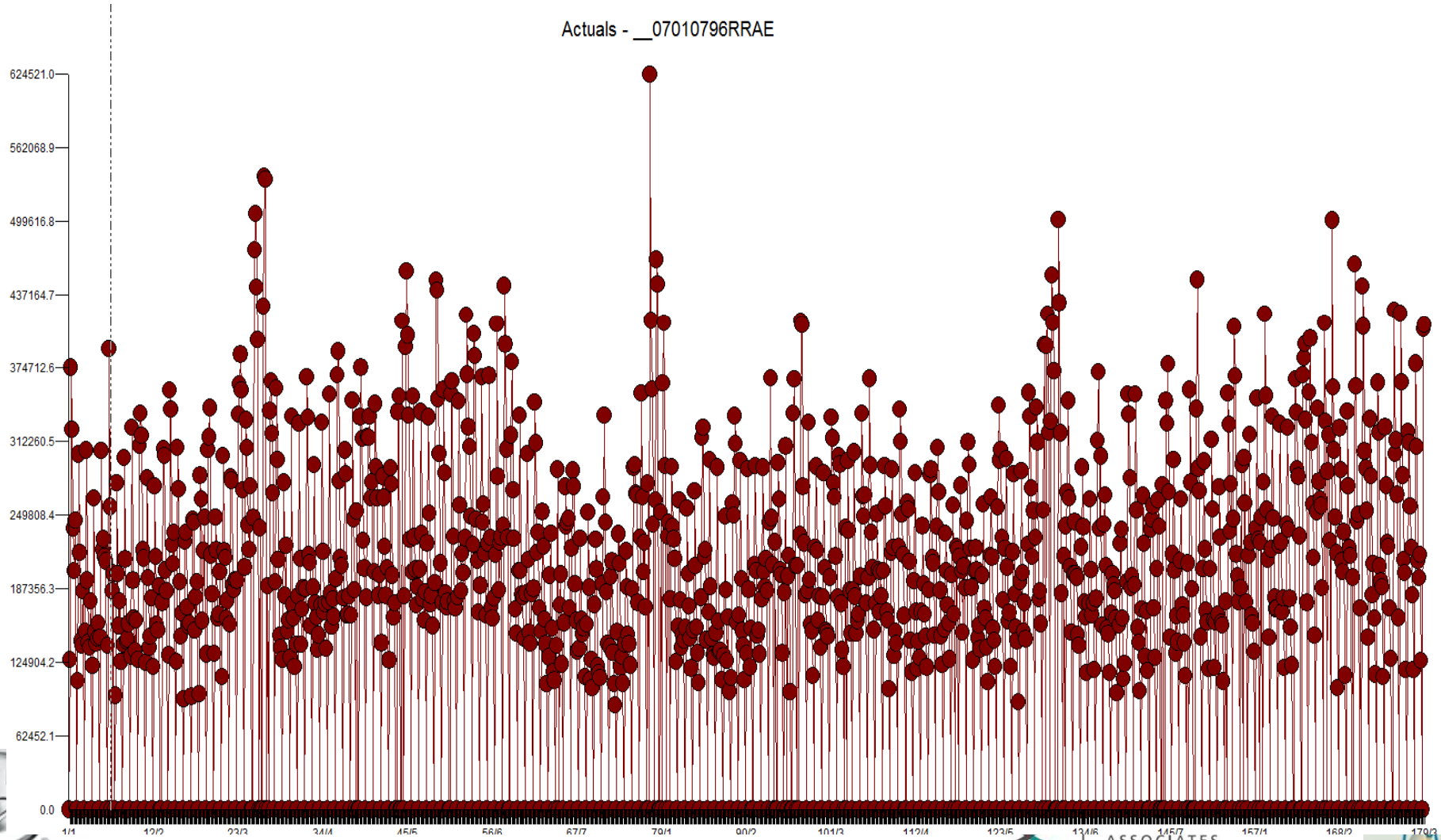
- Is the methodology explained? Does it do a "Pick Best" and ignore the assumptions?
- Does it build its own model from the information in each dataset or does it fit a couple of types of models to the data?
- Do they produce residuals that are free of pattern?
- Does the procedure explain what types of interventions it can detect?
- Does it explain what it does with interventions?
  - Does it correct for the outliers or just report them to you?
- Big ERP systems like Oracle, SAP, Manugistics, i2 are not exempt from this scrutiny

MERCURY MARINE

business modelling ASSOCIATES

AUTOMATIC FORECASTING SYSTEMS

3 ½ years of daily data – Cash management
Clear as mud?



MERCURY MARINE

ASSOCIATES

AUTOMATIC FORECASTING SYSTEMS

Distribution of Cash across all historical months

- Data begins on a Sunday, 7/1/2007, for daily cash demand
- Sundays are always 0
- There are many impacts on the data:
  - Trends
  - Seasonality
    - Monthly or Weekly patterns
  - Level
    - Big increases and drops, but not necessarily a trend
  - Autoregressive behavior
  - Day of the week
  - Fixed Day of the month
  - Seasonal Pulses - Changes in Day of the week
  - Interventions

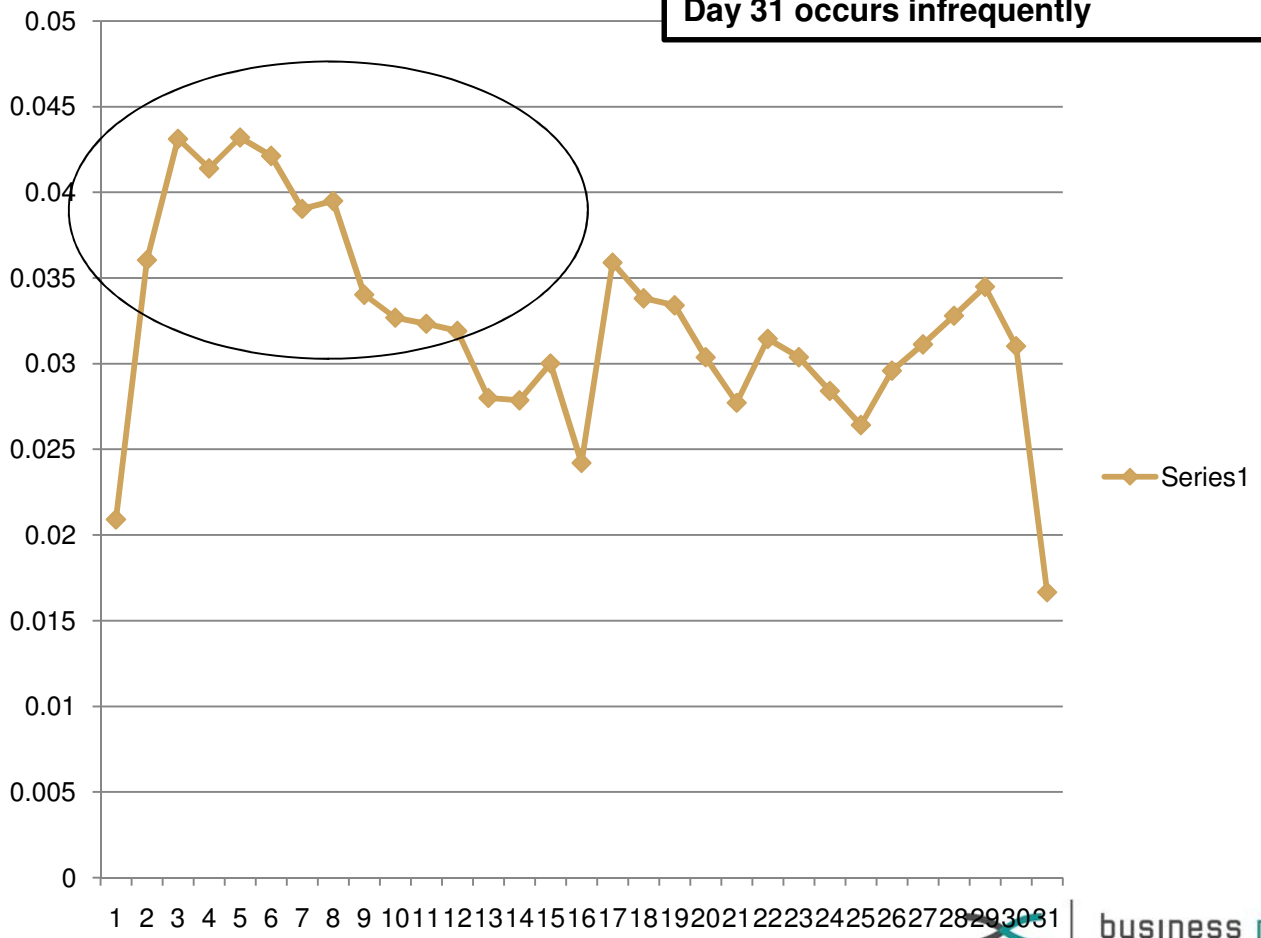
MERCURY MARINE

business modelling ASSOCIATES

AUTOMATIC FORECASTING SYSTEMS

Distribution of 3 ½ years of daily cash demand by day of month

Volume skewed towards the beginning of the month
Day 1 is impacted by New Year's and Worker's Day
Day 31 occurs infrequently

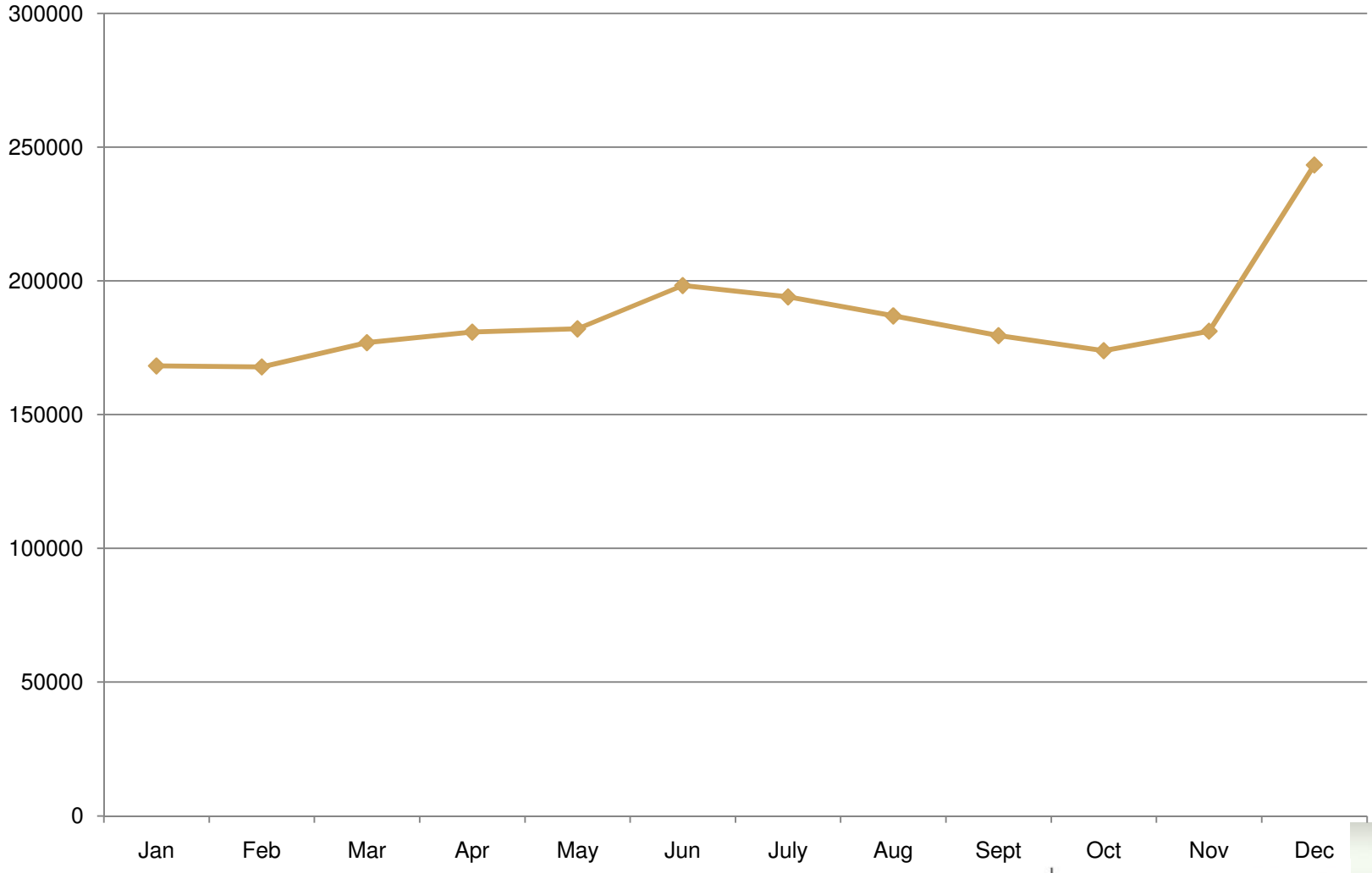


MERCURY MARINE

business modelling ASSOCIATES

AUTOMATIC FORECASTING SYSTEMS

Distribution of daily demand across 3 ½ years by month

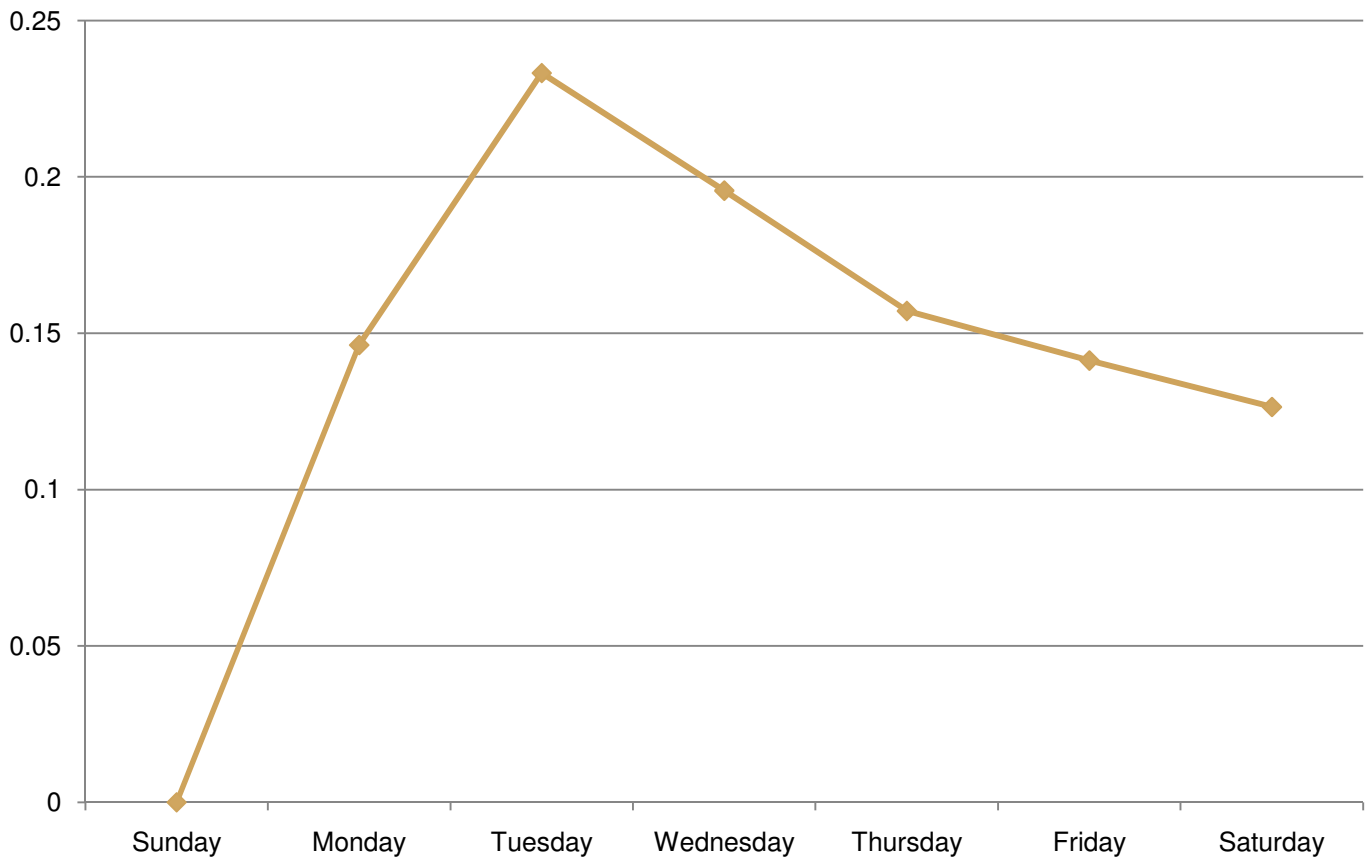


MERCURY MARINE

business modelling ASSOCIATES

AUTOMATIC FORECASTING SYSTEMS

Distribution of daily demand across 3 ½ years by day of week



MERCURY MARINE

business modelling ASSOCIATES

AUTOMATIC FORECASTING SYSTEMS

Variables in the model (partial list)

```
Estimation/Diagnostic Checking for Variable Y     __07010796RRAE
                                            X1    G_WOMEN
                                            X2    G_HERITAGE
                                            X3    G_RECONCILE
                                            X4    M_XMAS
                                            X5    M_NEWYEARS
                                            X6    M_EASTER
                                            X7    G_FREEDOM
                                            X8    G_WORKERS
                                            X9    G_YOUTH
                                            X10   MONTH_EFF10
                                            X11   MONTH_EFF12
                                            X12   MONTH_EFF01
                                            X13   MONTH_EFF02
                                            X14   FIXED_EFF_N10107
                                            X15   FIXED_EFF_N10307
                                            X16   FIXED_EFF_N10407
                                            X17   FIXED_EFF_N10507
                                            X18   FIXED_EFF_N10607
                :  VERY SPECIAL DAY VARIABLE   X19   FIXED_DAY02
                :  VERY SPECIAL DAY VARIABLE   X20   FIXED_DAY03
                :  VERY SPECIAL DAY VARIABLE   X21   FIXED_DAY04
                :  VERY SPECIAL DAY VARIABLE   X22   FIXED_DAY05
                :  VERY SPECIAL DAY VARIABLE   X23   FIXED_DAY06
                :  VERY SPECIAL DAY VARIABLE   X24   FIXED_DAY07
                :  VERY SPECIAL DAY VARIABLE   X25   FIXED_DAY08
                :  VERY SPECIAL DAY VARIABLE   X26   FIXED_DAY09
                :  NEWLY IDENTIFIED VARIABLE   X27   I~S00814 09/21/09    SEASP
                :  NEWLY IDENTIFIED VARIABLE   X28   I~P00170 12/17/07    PULSE
                :  NEWLY IDENTIFIED VARIABLE   X29   I~P01038 05/03/10    PULSE
```


MERCURY MARINE

business modelling ASSOCIATES

AUTOMATIC FORECASTING SYSTEMS

How complicated is this to do well?

#	MODEL COMPONENT	LAG (BOP)	COEFF	STANDARD ERROR	P VALUE	T VALUE
1	CONSTANT		.148E+06	.357E+04	.0000	41.36
	INPUT SERIES X1 G_WOMEN					
2	Omega (input) -Factor # 1	0	-.233E+06	.214E+05	.0000	-10.90
	INPUT SERIES X2 G_HERITAGE					
3	Omega (input) -Factor # 2	0	-.187E+06	.178E+05	.0000	-10.53
	INPUT SERIES X3 G_RECONCILE					
4	Omega (input) -Factor # 3	-1	.895E+05	.211E+05	.0000	4.24
5		0	.279E+06	.256E+05		
6		1	-.189E+06	.358E+05		
	INPUT SERIES X4 M_XMAS					
7	Omega (input) -Factor # 4	-4	.162E+06	.257E+05		6.32
8		-3	-.146E+06	[illegible]	.0000	-6.93
9		-2	-.162E+06	[illegible]	.0000	-6.31
10		-1	[illegible]	.211E+05	.0000	-6.29
11		0	.243E+06	.211E+05	.0000	11.53
12		1	.215E+06	.211E+05	.0000	10.19
	INPUT SERIES X5 M_NEWYEARS					
13	Omega (input) -Factor # 5	-4	.295E+06	.257E+05	.0000	11.50
14		-3	-.159E+06	.256E+05	.0000	-6.22
15		-2	-.476E+05	.211E+05	.0241	-2.26
16		0	.206E+06	.208E+05	.0000	9.92

Impacts from 4 days in advance of the holiday, on the holiday and the day after the holiday

MERCURY MARINE

business modelling ASSOCIATES

AUTOMATIC FORECASTING SYSTEMS

How complicated is this to do well?

```
INPUT SERIES X6  M_EASTER

 17Omega (input) -Factor #  6   -3  -.800E+05   .207E+05   .0001    -3.87
 18                             -2   .201E+06   .207E+05   .0000     9.70
 19                             -1  -.727E+05   .206E+05   .0004    -3.52
 20                              1   .210E+06   .206E+05   .0000    10.16

INPUT SERIES X7  G_FREEDOM

 21Omega (input) -Factor #  7    0  -.768E+05   .251E+05   .0023    -3.06
 22                              1  -.777E+05   .251E+05   .0020    -3.09

INPUT SERIES X8  G_WORKERS

 23Omega (input) -Factor #  8   -2   .769E+05   .205E+05   .0002     3.75
 24                              0   .174E+06   .206E+05   .0000     8.45

INPUT SERIES X9  G_YOUTH

 25Omega (input) -Factor #  9    0  -.238E+06   .205E+05   .0000   -11.63
 26                              1  -.657E+05   .205E+05   .0014    -3.20
 27                              2  -.757E+05   .205E+05   .0002    -3.69
```


MERCURY MARINE

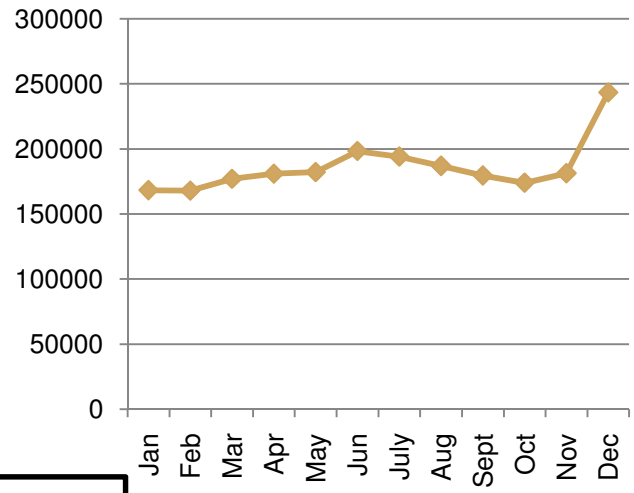
business modelling ASSOCIATES

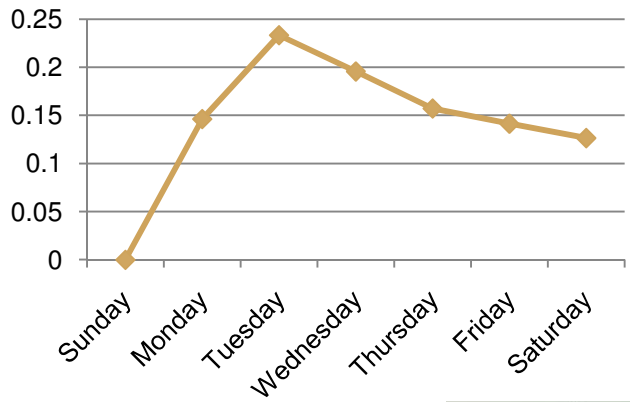
AUTOMATIC FORECASTING SYSTEMS

December is High, January, February and October Lower

```
INPUT SERIES X 10 MONTH_EFF10
 28Omega (input) -Factor # 10    0  -.121E+05   .350E+04   .0006    -3.44
INPUT SERIES X 11 MONTH_EFF12
 29Omega (input) -Factor # 11    0   .240E+05   .524E+04   .0000     4.58
INPUT SERIES X 12 MONTH_EFF01
 30Omega (input) -Factor # 12    0  -.128E+05   .404E+04   .0015    -3.18
INPUT SERIES X 13 MONTH_EFF02
 31Omega (input) -Factor # 13    0  -.205E+05   .416E+04   .0000    -4.93
INPUT SERIES X 14 FIXED_EFF_N10107
 32Omega (input) -Factor # 14    0  -.170E+06   .360E+04   .0000   -47.28
INPUT SERIES X 15 FIXED_EFF_N10307
 33Omega (input) -Factor # 15    0   .135E+06   .359E+04   .0000    37.51
INPUT SERIES X 16 FIXED_EFF_N10407
 34Omega (input) -Factor # 16    0   .836E+05   .363E+04   .0000    23.05
INPUT SERIES X 17 FIXED_EFF_N10507
 35Omega (input) -Factor # 17    0   .339E+05   .365E+04   .0000     9.30
INPUT SERIES X 18 FIXED_EFF_N10607
 36Omega (input) -Factor # 18    0   .190E+05   .363E+04   .0000     5.25
```

Remember Sunday is the first day of the series



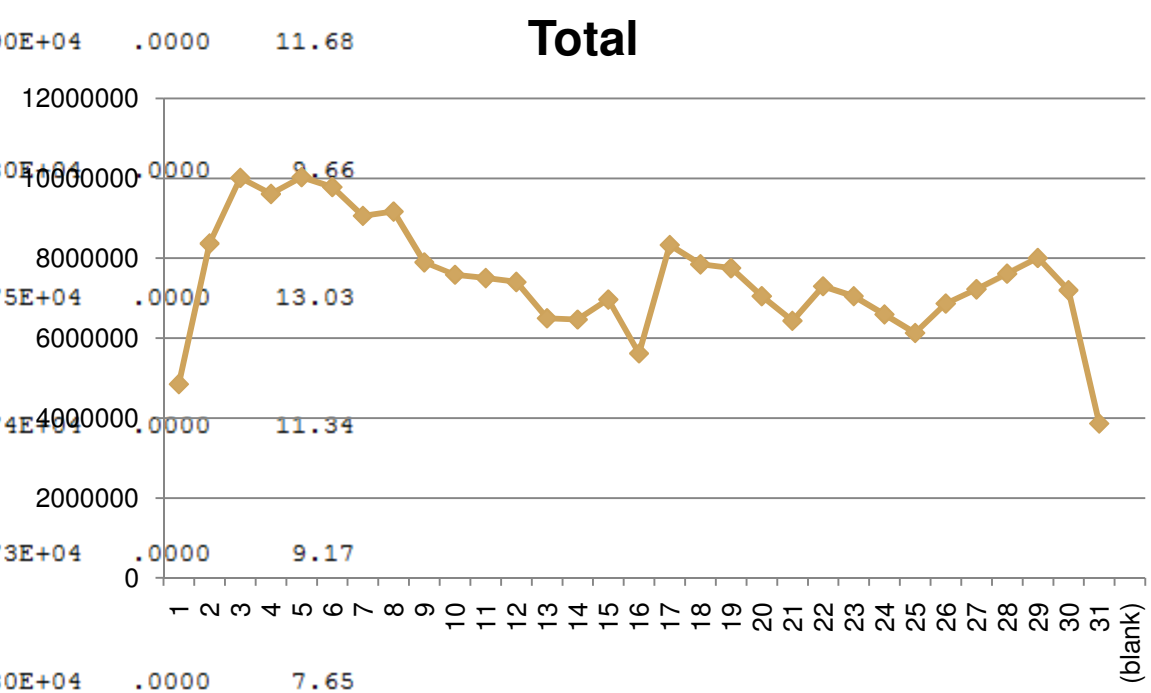


business modelling ASSOCIATES

AUTOMATIC FORECASTING SYSTEMS

Day 2 through 9 days are different than other days

```
INPUT SERIES X 19 FIXED_DAY02
 37Omega (input) -Factor # 19    0   .485E+05   .589E+04   .0000     8.22
INPUT SERIES X 20 FIXED_DAY03
 38Omega (input) -Factor # 20    0   .689E+05   .590E+04   .0000    11.68
INPUT SERIES X 21 FIXED_DAY04
 39Omega (input) -Factor # 21    0   .560E+05   .580E+04   .0000     9.66
INPUT SERIES X 22 FIXED_DAY05
 40Omega (input) -Factor # 22    0   .749E+05   .575E+04   .0000    13.03
INPUT SERIES X 23 FIXED_DAY06
 41Omega (input) -Factor # 23    0   .651E+05   .574E+04   .0000    11.34
INPUT SERIES X 24 FIXED_DAY07
 42Omega (input) -Factor # 24    0   .525E+05   .573E+04   .0000     9.17
INPUT SERIES X 25 FIXED_DAY08
 43Omega (input) -Factor # 25    0   .443E+05   .580E+04   .0000     7.65
INPUT SERIES X 26 FIXED_DAY09
 44Omega (input) -Factor # 26    0   .362E+05   .597E+04   .0000     6.05
```



MERCURY MARINE

business modelling ASSOCIATES

AUTOMATIC FORECASTING SYSTEMS

How complicated is this to do well?

Monday was not identified as a day of the week variable, but half way through it was found to have become different than the average

Multiple outliers that need to be cleansed in order to measure the true patterns

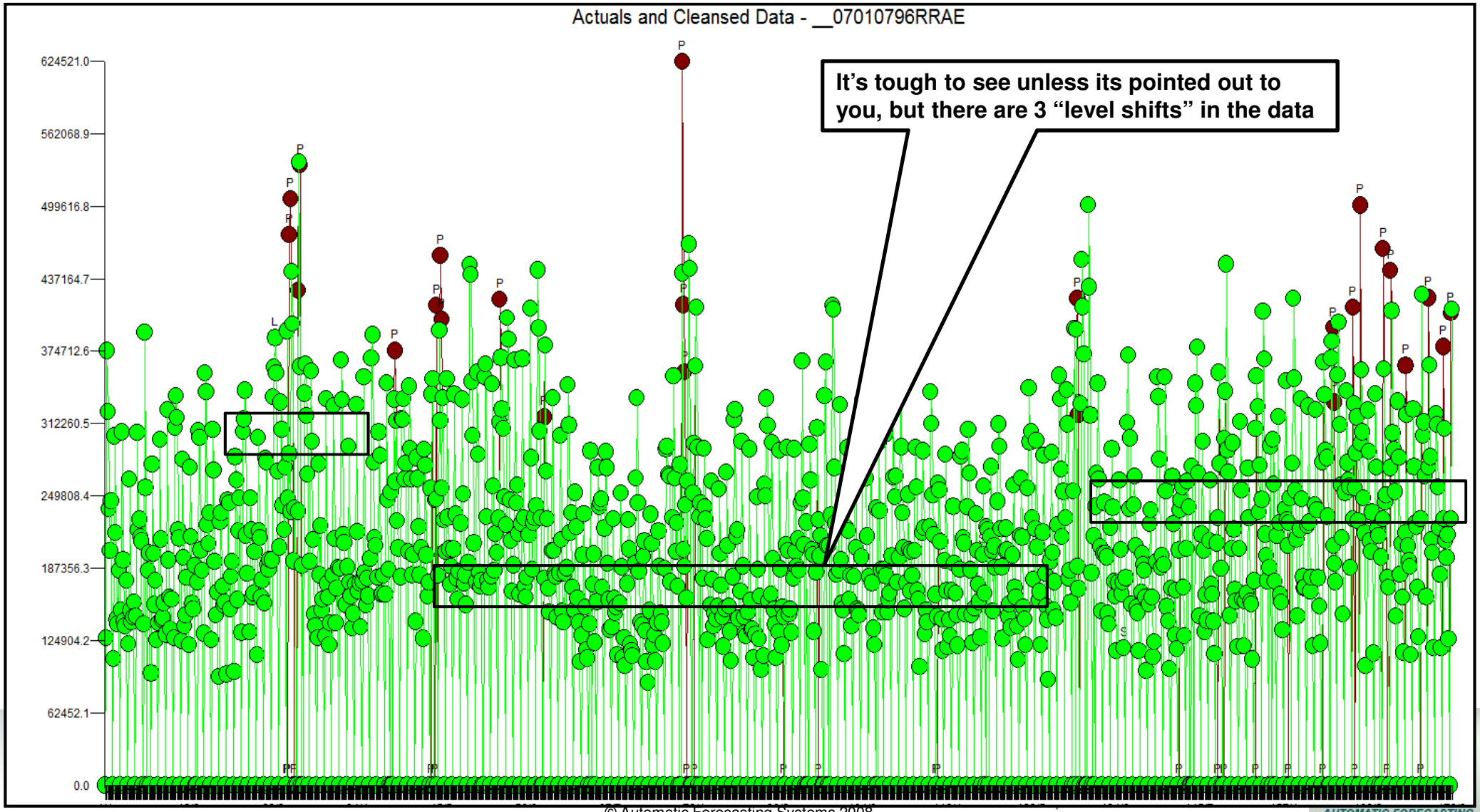
```
INPUT SERIES X 27 I~S00814 09/21/09    SEASP

 45Omega (input) -Factor # 27    0   .574E+05   .576E+04

INPUT SERIES X 28 I~P00170 12/17/07    PULSE

 46Omega (input) -Factor # 28    0  -.392E+06   .502E+05   .0000    -7.81

INPUT SERIES X 29 I~P01038 05/03/10    PULSE

 47Omega (input) -Factor # 29    0  -.299E+06   .362E+05   .0000    -8.25

INPUT SERIES X 30 I~L00998 03/24/10    LEVEL

 48Omega (input) -Factor # 30    0   .270E+05   .307E+04   .0000     8.80

INPUT SERIES X 31 I~L00417 08/20/08    LEVEL

 49Omega (input) -Factor # 31    0  -.328E+05   .277E+04   .0000   -11.83

INPUT SERIES X 32 I~L00159 12/06/07    LEVEL

 50Omega (input) -Factor # 32    0   .307E+05   .380E+04   .0000     8.06

INPUT SERIES X 33 I~P01067 06/01/10    PULSE

 51Omega (input) -Factor # 33    0  -.307E+06   .355E+05   .0000    -8.65
```


MERCURY MARINE

ness modelling ASSOCIATES

AUTOMATIC FORECASTING SYSTEMS

Actual(Red) and Cleansed(Green) History



Questions?

Automatic Forecasting Systems, Inc. (AFS)

P.O. Box 563

Hatboro, PA 19040

Phone: 215-675-0652

Fax: 215-672-2534

email: sales@Autobox.com

Web Site: www.Autobox.com