Capabilities of Autobox
Agenda

- Our Company & Awards
- Autobox Functionality
- Outliers will skew your model and forecast if not addressed
- Using Causal Variables
- Three Examples
- Questions
Our Company

- Incorporated in 1975
- First-to-market Forecasting package
- “AutoBJ” available in 1976 on Mainframe Time-sharing Services – IDC, CSC and Compuserve
- Autobox 1.0 launched DOS version on the PC in 1982
- Windows Version in 1991
- Batch Version 1996
- UNIX/AIX/SUN Version in 1999
- Callable DLL Version in 1999 for “plug and play” into ERP systems
- Java bean success in 2004
- .NET DLL version in 2013
- Delivering Price Elasticities using a Robust model 2014
- LLamasoft Integration 2014
- SIPmath Tools Certified 2015 – Delivering a distribution of forecasts for shortfall simulation “Simulated Forecasting
- Launched Autobox Integrated in R 2016
- Gartner Hype Cycle 2016 for our Probability Management Work delivering Simulated Forecasts
Awards

- Picked as the “Best Dedicated Forecasting” Software in the “Principles of Forecasting” textbook (Go to page 671 for overall results)

- Placed 12th in the “NN5” 2008 Forecasting Competition on “Daily data” (See www.neural-forecasting-competition.com results), but 1st among Automated software.

- Placed 2nd in the “NN3” 2007 Forecasting Competition on “Monthly data” (See www.neural-forecasting-competition.com), but 1st on more difficult data sets.
Specific Uses of Time Series Data Applied to any Industry
or What Can Autobox Be Used For?

- Data Cleansing - Correct historical data to what it should have been due to misreporting or removing the impact of unexpected events (i.e., outliers)

- Causation – Does my advertising (sentiment data) generate sales? Is Unemployment important? Evaluate historical data to determine if a variable is important and what is the exact time lag or lead? Outliers? Flagging a change in the model?

- Forecasting – Forecast incorporating future expected events
  - Short-Term (ATM) and Long-Term Demand of anything needing to be tracked and measured including regions
  - Daily Call Center Planning or Staffing at Departments (i.e., Hospital)
  - Intermittent Demand Data (Spare Parts with many zeroes in the data set)
  - Financial/Marketing - Probability of Hitting the Monthly/Quarterly #’s (Daily Data)
  - Capacity Planning or Risk of Insolvency

- What-if Analysis – Forecast using different scenarios to assess expected impact by changing future causal values (i.e., 1% increase in unemployment)

- Early Warning System/Detecting Change – Where are we underperforming/performing? or Fraud Detection Identify “most unusual” based on markers like outliers or changes in ---- trends/level/parameters/variance

- Price Elasticities with a robust model vs. the error prone log/log modeling with not outlier checking.

- Safety Stock for Inventory Control

- Simulated Forecasts – Certified by ProbabilityManagement.org for compatible SIPMath output – Providing more realistic confidence limits by way of sampling the errors, allowing outliers to “play”
Recent Advancements

- Univariate and Multivariate (Two-stage Monte Carlo) Simulation
- Resample the Errors
- Model and account for the outliers, but then allow outliers to “play” (ie exist) in the forecast
- Confidence Limits are too tight!!
- Symmetrical Confidence Limits → Asymmetrical Confidence Limits
- Averages aren’t added, but Distributions are. A 1,000 forecasts from Autobox can be integrated into Probability Management’s Capacity/Shortfall model
- User supplied estimates of expectations (ie Delphi forecast) with probabilities and ranges can be included as a causal
Capabilities of Autobox
Some Customers

- National Defence
- PRAXAIR
- monroe plan
- Schweizerische Eidgenossenschaft
- Confederazione Svizzera
- Confederazione svizra
- TEXAS Department of Assistive and Rehabilitative Services
- BRIDGESTONE
Autobox has been used in articles published in a variety of Journals as it has unique strengths not found in other software. Read the articles in our ‘News” section on the website.

- Journal of Forecasting
- Journal of Business Forecasting
- North American Actuarial Journal
- Forest Research and Management Institute
- Environmental and Resource Economics
- Technological Forecasting and Social Change
- Fraud Magazine
- Canadian Journal of Forest Research
- Applied Economics
- Journal of applied Pharmacology
- Journal of Endocrinology and Metabolism
- Journal of Urban Studies
Autobox Functionality
What kinds of Products do you Offer?

There are a couple of ways to engage Autobox

- **Autobox Interactive** – Run one series at a time in a Windows environment

- **Autobox Batch**
  - Interactive and DOS – Read from Excel to run Univariate problems and post results to Excel
  - Interactive – Read flat files for Univariate and Multivariate problems
  - Command Line – Read flat file for Univariate and Multivariate problems – Parallel processing as it can run up to as many CPUs on your machine

- **Autobox for R** – callable within R (not free 😔)

- **Integrated** - Call the Autobox DLL directly from your application
Statistical Modeling Sample Space
Autobox vs i2 (You can substitute any “pick best“ here)
Autobox - Process

- Data
- Identifying Relationships (i.e., identify the signal from the noise)
- Tweaking the model (i.e., identifying the unusual in the signal)

Forecast
## Autobox Functionality

### Data
- Historical Data
- Historical Causal Data (e.g., Price, Customer Insight, GNP, Unemployment, Population, etc.)
- Historical Knowledge of Events (outliers, mergers, promotions, holidays, etc.)
- Future Values of Causals
- Future Values of Promotions

### Identifying Relationship
- Autobox Algorithm *(Can be customized by user)*
- Autocorrelation Function
- Partial Autocorrelation Function
- Cross Correlation Function
- ARIMA models
- Transfer Function models

### Tweaking the model
- Autobox will identify the lead/lag relationship in the causals
- Autobox will remove unimportant Causals
- Autobox will potentially identify and add 4 Types of Outliers providing “Early Warning Detection”
- Check for Constancy of Parameters and Variance

### Forecast
- What if scenarios to evaluate different scenarios using different future values of causals

---

© Automatic Forecasting Systems 2019
Forecasting Methods Family Tree

**Forecasts**

**Judgmental**
- Field Sales - Business partners
- Jury of Executives
- Delphi

**Time Series - Unknown causals**

**Smoothing**

**Decomposition**

**Autocorrelation correction - ARIMA models**

**Moving Averages**

**Simple Regression**

**Multiple Regression**

**ARIMA transfer function models**

**Causal**

- Integrated time series and causal
  - Single Exponential Smoothing
  - Double (Holt’s) Exponential Smoothing
  - Triple (Winters’) Exponential Smoothing
Regression Modelling

- Causal variables - Variables like Price, Unemployment, Population might have a lead or lag relationship and that exact period may be difficult to identify.

- Dummy variables – Outliers like Promotions, Earthquakes, mergers should be provided and/or identified and adjusted for by the system.

- Memory – There is a period to period relationship that exists in the data like this month to last month and to last year.
Autobox – It’s information Rich – White box not Black Box

- Reports on
  - Analytical Steps, Interventions, Equations and Overall Summary
  - Early Warning System showing series with unusual values in the latest data period and Pulse Report showing outliers at similar periods
  - Forecasts, Forecasts of causals if no forecast exists, Cleansed historical data, Safety Stock for 90% service level

© Automatic Forecasting Systems 2019
Why is Autobox’s Methodology Different?

- **Automatically** creates a customized model for every data set. Not “pick best”
- **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
How Autobox Treats Different Data Intervals
You can (optionally) let the system do it all by itself!

• Incorporates variables for Hourly Data - Brings in Daily History and Forecast as a Causal Variable for the 24 separate regressions

• Incorporates variables for Daily data **Automatically**:
  • Day of the week (i.e. Sundays are low)
  • Special Days of the month (i.e. Social Security checks are delivered on the 3rd of the month)
  • Week of the Year (i.e. 51 dummies Capturing seasonal variations) or Month of the Year (i.e. 11 dummies Capturing seasonal variations)
  • Adds in holiday variables (including “Fridays before" holidays that fall on Mondays and Monday after a Friday Holiday AND a separate effect “long weekends”)
  • End of the Month Effect – when last day of month is a Friday, Saturday or Sunday

• Incorporates variables for Weekly data:
  • Trading Days (i.e. 19,19,22,20,21,21, etc.)
  • Week of the Year (i.e. Capturing seasonal variations) **Automatically**

• Incorporates variables for Monthly data:
  • Trading Days (i.e. 19,19,22,20,21,21, etc.)
  • Month of the Year (i.e. 11 dummies Capturing seasonal variations)
  • Accounting effect (i.e. 4/4/5)
    • Accounting practice of uneven grouping of weeks into monthly buckets where there is a 4/4/5 pattern that is repeated

• Incorporates variables for Quarterly data:
  • Quarterly effect (i.e. High in Q2)
Outliers will skew your model and forecast if not addressed

Time Series Analysis ASSUMES that the errors are N.I.I.D. with a constant mean and constant variance. Most ignore this assumption and don’t even provide graphs to show if it is or isn’t!
Outlier Detection – Pulse(s)  
(FIXING THE HISTORICAL DATA)

- Pulse – Fire in the warehouse in April  (0,0,0,0,0,0,0,0,1,0,0,0,0,0)
Outlier Detection – Seasonal Pulse(s)  
(ADAPTING TO THE DATA)

- Seasonal Pulse – February emerges later during the year (0,1,0,0,0,0,0,0,0,0,0,0,1)
Outlier Detection – Level Shift(s)
(ADAPTING TO THE DATA)

- 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). Not a trend!
Outlier Detection – Local Time Trend(s)
(ADAPTING TO THE DATA)

- 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.)
Outlier Detection – Local Time Trend(s)
(ADAPTING TO THE DATA)

- 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.)
Parameter Change Detection – Chow Test
(DELETING OLDER DATA)

Gregory Chow applied an approach and F Test to determine if two models were similar. We apply this to time series and find the model has changed.

The program has found a statistically significant difference between older data and more recent data. Autobox reported a structural break in the parameters starting at 1998/3. Thus 75 observations have been set aside and not used in the final model. The most recent 48 observations starting at 1998/3 were the basis for the final model/parameters.
Parameter Change Detection – Chow Test
(DELETING OLDER DATA)

- Older data is truncated and not used
Variance Change Detection – Tsay Test

(WEIGHTING DATA)

- Weighted Least Squares
Variance Change Detection – Tsay Test (WEIGHTING DATA)

- Weighted Least Squares

**Standardized Values are indicated** \( \frac{(X - \text{Mean})}{\text{SIGMA}} \)
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 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.
What is unusual?

- We see a big outlier, but what about the pattern near the end?
- Do we remove/fix those also? Do we adjust them to be an average of the previous data points?
Was it a Causal Model Issue all along?

- 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 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

© Automatic Forecasting Systems 2019
What is unusual?

- This is an 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 inlier 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.
Here is an outlier, right?

- This value is **not** an unusual data point

```plaintext
Actuals - VC0699NT46
```

© Automatic Forecasting Systems 2019
Seasonal Pulses

- The outlier is really a series of outliers called a seasonal pulse
Autobox and Inliers

- This is a dramatic example, but it illustrates the power of the methodology.
- Can your software detect an outlier at the mean?
- Or Does it just use 2 sigma around the mean and hope for the best?
- In order to detect what is unusual, you need to detect what is usual.
- This is why we create a model and not simply a force an existing model to data.
Outlier Detection – What should you do about it?

- User Provides knowledge (APRIORI) 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 (i.e., Intervention modeling) as a possible input variable. In this case, an actual variable now has a coefficient and can be used to explain the impact (lift or decrease) which will yield a better model and forecast.

- 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 bad model and forecast.
Early Warning System Reports Tells you the Probability of the last observation being “out of control” Statistically

- You can inform senior management which SKU's seem to be out of control based on the latest observation.
- Instead of using an arbitrary measure to detect unusual behavior (i.e. % change from last year or % difference using the last two periods), Autobox tells you “the probability of observing that last observation before it was observed”.
- Autobox will test and report on the probability that the last observation is unusual. It will write out a report for every series analyzed which can then be sorted to identify those series that look to be unusual.
- Here we run the series “inlier” and the report shows no warning in the “probability” field when the actual last value was 9.0. When we change that value to a 5.0 and rerun then AUTOBOX reacts and the small P-value reported showing us significance.
Autobox’s Pulse/Level/Seasonal Pulse Trend/Variance/Parameter Early Warning Reports

- Autobox reports all outliers across all time periods so you can easily identify out of control behavior from a macro level suggesting a widespread event (known or possibly unknown) for research or just understanding.

- Just import the report into Excel, sum the columns and transpose to identify time periods with multiple pulses in the same time period. 2 of the 3 SKUs show period 30, 41, and 52 with an intervention. This might spark some discussion as to why this is occurring. It may be random or part of a systematic event. If so, then a causal variable could be introduced into Autobox to “model” its effect and plan for the impact in the future.

<table>
<thead>
<tr>
<th>SERIES</th>
<th>#</th>
<th>NOB</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>4</td>
<td>54</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b</td>
<td>2</td>
<td>54</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c</td>
<td>4</td>
<td>54</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>time period</th>
<th>count</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>2</td>
</tr>
<tr>
<td>41</td>
<td>2</td>
</tr>
<tr>
<td>52</td>
<td>2</td>
</tr>
<tr>
<td>23</td>
<td>1</td>
</tr>
<tr>
<td>26</td>
<td>1</td>
</tr>
<tr>
<td>28</td>
<td>1</td>
</tr>
<tr>
<td>31</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
</tr>
</tbody>
</table>

© Automatic Forecasting Systems 2008
Autobox’s Forecast Reconciliation

- Top level and SKU level forecasts are reconciled in two ways:
  - “Top-down” reconciliation
    - Create Forecast for the Top level
    - Force the bottom level to match the top level using an allocation of the forecast period by period
  - “Bottom-Up” with no reconciliation
    - Create Forecast for the Bottom level
    - Aggregate the Bottom level to be the Top level forecast
Using Causal Variables
Two Types of Users
Rear View Mirror vs. Rear and Front Windshield

- Use the History of the data only

- Use the History of the data AND causal variables (i.e. holidays, price, marketing promotions, advertising) and the future values of these variables.
Case Study – What-if Analysis

- Client wanted a national model using the ability to incorporate causal variables and create scenarios using different levels of causal variables using weekly data. Here are the causals:
  - Average unit Price
  - Total number of stores
  - Marketing Index
  - Holiday variables
  - TV GRPs
Case Study – What-if Analysis
Baseline Forecast

Actuals and Forecasts - SALES

Periods 2000/27 to 2003/21 (Seasonality of 52)
Case Study – What-if Analysis
Baseline Future Values of Causals

<table>
<thead>
<tr>
<th>BY/EP</th>
<th>ORIGINAL FCSTs</th>
<th>ORIGINAL FVs</th>
<th>ORIGINAL FVs</th>
<th>ORIGINAL FVs</th>
<th>ORIGINAL FVs</th>
<th>ORIGINAL FVs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SALES</td>
<td>AUP</td>
<td>QM</td>
<td>TD</td>
<td>TPR</td>
<td>TV</td>
</tr>
<tr>
<td>2003/5</td>
<td>62043.51445260</td>
<td>2.94000000</td>
<td>6.00000000</td>
<td>290.00000000</td>
<td>12.00000000</td>
<td>11.00000000</td>
</tr>
<tr>
<td>2003/6</td>
<td>59723.17764694</td>
<td>2.84000000</td>
<td>6.00000000</td>
<td>290.00000000</td>
<td>12.00000000</td>
<td>9.00000000</td>
</tr>
<tr>
<td>2003/7</td>
<td>605600.605367</td>
<td>2.94000000</td>
<td>6.00000000</td>
<td>290.00000000</td>
<td>12.00000000</td>
<td>7.00000000</td>
</tr>
<tr>
<td>2003/8</td>
<td>77341.43308286</td>
<td>2.84000000</td>
<td>6.00000000</td>
<td>290.00000000</td>
<td>12.00000000</td>
<td>13.00000000</td>
</tr>
<tr>
<td>2003/9</td>
<td>606557.2665496</td>
<td>2.96000000</td>
<td>11.00000000</td>
<td>280.00000000</td>
<td>15.00000000</td>
<td>18.00000000</td>
</tr>
<tr>
<td>2003/10</td>
<td>621441.2477218</td>
<td>2.86000000</td>
<td>11.00000000</td>
<td>288.00000000</td>
<td>15.30000000</td>
<td>21.00000000</td>
</tr>
<tr>
<td>2003/11</td>
<td>640511.2200540</td>
<td>2.96000000</td>
<td>11.00000000</td>
<td>280.00000000</td>
<td>15.00000000</td>
<td>17.00000000</td>
</tr>
<tr>
<td>2003/12</td>
<td>615173.2625224</td>
<td>2.86000000</td>
<td>11.00000000</td>
<td>288.00000000</td>
<td>15.30000000</td>
<td>21.00000000</td>
</tr>
<tr>
<td>2003/13</td>
<td>634167.3511666</td>
<td>2.96000000</td>
<td>11.00000000</td>
<td>288.00000000</td>
<td>15.30000000</td>
<td>24.00000000</td>
</tr>
<tr>
<td>2003/16</td>
<td>642579.6665278</td>
<td>2.93000000</td>
<td>4.00000000</td>
<td>285.00000000</td>
<td>10.00000000</td>
<td>25.00000000</td>
</tr>
<tr>
<td>2003/17</td>
<td>623012.85383935</td>
<td>2.93000000</td>
<td>4.00000000</td>
<td>285.00000000</td>
<td>10.00000000</td>
<td>29.00000000</td>
</tr>
<tr>
<td>2003/18</td>
<td>552651.9382621</td>
<td>2.94000000</td>
<td>4.00000000</td>
<td>283.00000000</td>
<td>4.00000000</td>
<td>18.00000000</td>
</tr>
<tr>
<td>2003/19</td>
<td>533357.5281302</td>
<td>2.94000000</td>
<td>4.00000000</td>
<td>283.00000000</td>
<td>4.00000000</td>
<td>13.00000000</td>
</tr>
<tr>
<td>2003/20</td>
<td>536594.3015699</td>
<td>2.94000000</td>
<td>4.00000000</td>
<td>283.00000000</td>
<td>4.00000000</td>
<td>10.00000000</td>
</tr>
<tr>
<td>2003/21</td>
<td>511661.8726510</td>
<td>2.94000000</td>
<td>4.00000000</td>
<td>283.00000000</td>
<td>4.00000000</td>
<td>8.00000000</td>
</tr>
</tbody>
</table>

SUM: 10058442, 9378767
Case Study – What-if Analysis
Scenario #1 Adjust Price and TV Spots Up

<table>
<thead>
<tr>
<th>BY/YP</th>
<th>AUP</th>
<th>QM</th>
<th>TD</th>
<th>TFP</th>
<th>TV</th>
</tr>
</thead>
<tbody>
<tr>
<td>2003/5</td>
<td>2.6400000</td>
<td>0.0000000</td>
<td>290.0000000</td>
<td>12.0000000</td>
<td>11.0000000</td>
</tr>
<tr>
<td>2003/6</td>
<td>2.5300000</td>
<td>0.0000000</td>
<td>290.0000000</td>
<td>12.0000000</td>
<td>12.0000000</td>
</tr>
<tr>
<td>2003/7</td>
<td>2.9400000</td>
<td>0.0000000</td>
<td>290.0000000</td>
<td>12.0000000</td>
<td>12.0000000</td>
</tr>
<tr>
<td>2003/8</td>
<td>2.9400000</td>
<td>11.0000000</td>
<td>288.0000000</td>
<td>15.0000000</td>
<td>18.0000000</td>
</tr>
<tr>
<td>2003/9</td>
<td>2.9300000</td>
<td>11.0000000</td>
<td>288.0000000</td>
<td>15.0000000</td>
<td>21.0000000</td>
</tr>
<tr>
<td>2003/10</td>
<td>2.9300000</td>
<td>11.0000000</td>
<td>288.0000000</td>
<td>15.0000000</td>
<td>21.0000000</td>
</tr>
<tr>
<td>2003/11</td>
<td>2.9300000</td>
<td>11.0000000</td>
<td>288.0000000</td>
<td>15.0000000</td>
<td>21.0000000</td>
</tr>
<tr>
<td>2003/12</td>
<td>2.9300000</td>
<td>11.0000000</td>
<td>288.0000000</td>
<td>15.0000000</td>
<td>21.0000000</td>
</tr>
<tr>
<td>2003/14</td>
<td>2.9300000</td>
<td>4.0000000</td>
<td>265.0000000</td>
<td>10.0000000</td>
<td>15.0000000</td>
</tr>
<tr>
<td>2003/15</td>
<td>2.9300000</td>
<td>4.0000000</td>
<td>265.0000000</td>
<td>10.0000000</td>
<td>22.0000000</td>
</tr>
<tr>
<td>2003/16</td>
<td>2.9300000</td>
<td>4.0000000</td>
<td>265.0000000</td>
<td>10.0000000</td>
<td>25.0000000</td>
</tr>
<tr>
<td>2003/17</td>
<td>2.9300000</td>
<td>4.0000000</td>
<td>265.0000000</td>
<td>10.0000000</td>
<td>20.0000000</td>
</tr>
<tr>
<td>2003/18</td>
<td>2.9300000</td>
<td>4.0000000</td>
<td>265.0000000</td>
<td>10.0000000</td>
<td>16.0000000</td>
</tr>
<tr>
<td>2003/19</td>
<td>2.9300000</td>
<td>4.0000000</td>
<td>265.0000000</td>
<td>10.0000000</td>
<td>13.0000000</td>
</tr>
<tr>
<td>2003/20</td>
<td>2.9300000</td>
<td>4.0000000</td>
<td>265.0000000</td>
<td>10.0000000</td>
<td>10.0000000</td>
</tr>
<tr>
<td>2003/21</td>
<td>2.9300000</td>
<td>4.0000000</td>
<td>265.0000000</td>
<td>10.0000000</td>
<td>8.0000000</td>
</tr>
</tbody>
</table>
Case Study – What-if Analysis
Graph of Baseline and Scenario #1

Forecasts - All Whatif Scenarios - SALES

Periods 2003/5 to 2003/21 (Seasonality of 52)
Three Examples
Financial Forecasting Example
How is your Finance Team doing this now?

What's the probability of making the month end targeted number given the most recent daily observation?

- The 2008 financial crises caught a few companies unable to quickly identify when month end numbers were not going to be met.

- Simplistic approaches use a ratio estimate (ie 5 days into the month 30/5 so multiply current month total by 6 to get month end estimate) are simplistic and incorrect. Promotions and day of the week effects are not considered using ratio estimates and need to be modeled at a DAILY level as part of a comprehensive model and forecast which can then be used to determine probabilities of making the month end number.

- Autobox reports out a variety of Probabilities which the target can be evaluated against. A summary report can then be used to identify which SKU’s are likely NOT to make the month end number.
Graph of 3 ½ years of daily data – Clear as mud?
Daily Demand

• Data begins on a Sunday, 7/1/2007
• 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
  • Holidays plus before and after
Distribution of 3 ½ years of daily 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
Distribution of daily demand across 3 ½ years by month
Distribution of daily demand across 3 ½ years by day of week
Variables in the model (partial list)

Estimation/Diagnostic Checking for Variable Y _07010796RAE
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
How complicated is this to do well?

<table>
<thead>
<tr>
<th>MODEL COMPONENT</th>
<th>COEFF</th>
<th>STANDARD ERROR</th>
<th>P VALUE</th>
<th>T VALUE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1CONSTANT</td>
<td>.148E+06</td>
<td>.357E+04</td>
<td>.0000</td>
<td>41.36</td>
</tr>
<tr>
<td>INPUT SERIES X1 G_WOMEN</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2Omega (input)</td>
<td>-Factor # 1</td>
<td>0</td>
<td>-.233E+06</td>
<td>.214E+05</td>
</tr>
<tr>
<td>INPUT SERIES X2 G heritAGE</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3Omega (input)</td>
<td>-Factor # 2</td>
<td>0</td>
<td>-.187E+06</td>
<td>.178E+05</td>
</tr>
<tr>
<td>INPUT SERIES X3 G_RECONCILE</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4Omega (input)</td>
<td>-Factor # 3</td>
<td>-1</td>
<td>.895E+05</td>
<td>.211E+05</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>.279E+06</td>
<td>.256E+05</td>
<td>.0000</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>-.189E+06</td>
<td>.358E+05</td>
<td>.0000</td>
</tr>
<tr>
<td>INPUT SERIES X4 M_XMAS</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7Omega (input)</td>
<td>-Factor # 4</td>
<td>-4</td>
<td>.162E+06</td>
<td>.257E+05</td>
</tr>
<tr>
<td>8</td>
<td>-3</td>
<td>-.146E+06</td>
<td>.211E+05</td>
<td>.0000</td>
</tr>
<tr>
<td>9</td>
<td>-2</td>
<td>-.162E+06</td>
<td>.256E+05</td>
<td>.0000</td>
</tr>
<tr>
<td>10</td>
<td>-1</td>
<td>.208E+06</td>
<td>.211E+05</td>
<td>.0000</td>
</tr>
<tr>
<td>11</td>
<td>0</td>
<td>.243E+06</td>
<td>.211E+05</td>
<td>.0000</td>
</tr>
<tr>
<td>12</td>
<td>1</td>
<td>.215E+06</td>
<td>.211E+05</td>
<td>.0000</td>
</tr>
<tr>
<td>INPUT SERIES X5 M_NEWYEARS</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13Omega (input)</td>
<td>-Factor # 5</td>
<td>-4</td>
<td>.295E+06</td>
<td>.257E+05</td>
</tr>
<tr>
<td>14</td>
<td>-3</td>
<td>-.159E+06</td>
<td>.256E+05</td>
<td>.0000</td>
</tr>
<tr>
<td>15</td>
<td>-2</td>
<td>-.476E+05</td>
<td>.211E+05</td>
<td>.0241</td>
</tr>
<tr>
<td>16</td>
<td>0</td>
<td>.206E+06</td>
<td>.208E+05</td>
<td>.0000</td>
</tr>
</tbody>
</table>

Impacts from 4 days in advance of the holiday, on the holiday and the day after the holiday.
How complicated is this to do well?

<table>
<thead>
<tr>
<th>INPUT SERIES X6</th>
<th>M_EASTER</th>
</tr>
</thead>
<tbody>
<tr>
<td>17Omega (input)</td>
<td>-Factor # 6</td>
</tr>
<tr>
<td>18</td>
<td>-3</td>
</tr>
<tr>
<td>19</td>
<td>-2</td>
</tr>
<tr>
<td>20</td>
<td>-1</td>
</tr>
<tr>
<td>21</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>INPUT SERIES X7</th>
<th>G_FREEDOM</th>
</tr>
</thead>
<tbody>
<tr>
<td>21Omega (input)</td>
<td>-Factor # 7</td>
</tr>
<tr>
<td>22</td>
<td>0</td>
</tr>
<tr>
<td>23</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>INPUT SERIES X8</th>
<th>G_WORKERS</th>
</tr>
</thead>
<tbody>
<tr>
<td>23Omega (input)</td>
<td>-Factor # 8</td>
</tr>
<tr>
<td>24</td>
<td>-2</td>
</tr>
<tr>
<td>25</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>INPUT SERIES X9</th>
<th>G_YOUTH</th>
</tr>
</thead>
<tbody>
<tr>
<td>25Omega (input)</td>
<td>-Factor # 9</td>
</tr>
<tr>
<td>26</td>
<td>0</td>
</tr>
<tr>
<td>27</td>
<td>1</td>
</tr>
<tr>
<td>28</td>
<td>2</td>
</tr>
</tbody>
</table>
The model matches the data - 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
Day 2 through 9 are higher than other days

INPUT SERIES X 19 FIXED_DAY02
37Omega (input) - Factor # 19 0 0.485E+05 0.589E+04 0.0000 8.22

INPUT SERIES X 20 FIXED_DAY03
38Omega (input) - Factor # 20 0 0.689E+05 0.590E+04 0.0000 11.68

INPUT SERIES X 21 FIXED_DAY04
39Omega (input) - Factor # 21 0 0.560E+05 0.580E+04 0.0000 10.0

INPUT SERIES X 22 FIXED_DAY05
40Omega (input) - Factor # 22 0 0.749E+05 0.575E+04 0.0000 14.9

INPUT SERIES X 23 FIXED_DAY06
41Omega (input) - Factor # 23 0 0.651E+05 0.574E+04 0.0000 9.1

INPUT SERIES X 24 FIXED_DAY07
42Omega (input) - Factor # 24 0 0.525E+05 0.573E+04 0.0000 17.6

INPUT SERIES X 25 FIXED_DAY08
43Omega (input) - Factor # 25 0 0.443E+05 0.580E+04 0.0000 7.65

INPUT SERIES X 26 FIXED_DAY09
44Omega (input) - Factor # 26 0 0.362E+05 0.597E+04 0.0000 6.05

© Automatic Forecasting Systems 2019
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.

The “average” volume had temporary changes 3 times over the 3 ½ years.

<table>
<thead>
<tr>
<th>INPUT SERIES X 27</th>
<th>09/21/09</th>
<th>SEASP</th>
</tr>
</thead>
<tbody>
<tr>
<td>45Omega (input)</td>
<td>Factor #27</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>INPUT SERIES X 28</th>
<th>12/17/07</th>
<th>PULSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>46Omega (input)</td>
<td>Factor #28</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>INPUT SERIES X 29</th>
<th>05/03/10</th>
<th>PULSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>47Omega (input)</td>
<td>Factor #29</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>INPUT SERIES X 30</th>
<th>03/24/10</th>
<th>LEVEL</th>
</tr>
</thead>
<tbody>
<tr>
<td>48Omega (input)</td>
<td>Factor #30</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>INPUT SERIES X 31</th>
<th>08/20/08</th>
<th>LEVEL</th>
</tr>
</thead>
<tbody>
<tr>
<td>49Omega (input)</td>
<td>Factor #31</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>INPUT SERIES X 32</th>
<th>12/06/07</th>
<th>LEVEL</th>
</tr>
</thead>
<tbody>
<tr>
<td>50Omega (input)</td>
<td>Factor #32</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>INPUT SERIES X 33</th>
<th>06/01/10</th>
<th>PULSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>51Omega (input)</td>
<td>Factor #33</td>
<td>0</td>
</tr>
</tbody>
</table>
Actual(Red) and Cleansed(Green) History – Clear now!

It’s tough to see unless its pointed out to you, but there are 3 “level shifts” in the data.
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”
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.

![Standard Deviation vs Mean Unadjusted for Outliers](chart1)

![Standard Deviation versus Mean – Adjusted for Outliers](chart2)
Did you spot the outliers in 1960?

The last July is significantly higher than August.

October was unusually high.

The forecasts are not overly impacted by the anomalous values.

March is always a breakout month, but not here.
The Standard Forecasting Tools in APO

- Moving Averages and weighted moving averages
- A portion of the family of exponential smoothing methods (a notable exclusion being the set of procedures that assume multiplicative seasonality)
- Automatic model selection in which the system chooses among included members of the exponential smoothing family
- Croston’s model for intermittent demand (without the Syntetos and Boylan(2005) corrections)
- Simple and multiple regressions
Summary

SAP APO is focusing on the whole supply chain and also on planning and process consistency. The mathematical accuracy of its forecasts may be worse than that of a stand-alone forecasting package, but the benefits to our company in terms of worldwide network planning and control more than compensate for this.
Integrating Autobox into SAP
Writing to Disk

SAP

Sales History & Causal Variables Written to Disk

Autobox Batch EXE

Forecasts Written to Disk

SAP

© Automatic Forecasting Systems 2019
Integrating Autobox into SAP
Calling a DLL and What-if Scenario

Forecasts Passed to Autobox in memory

Forecasts Written or Passed to an application to do What-if Scenarios
Make sure to join the Autobox discussion group on linkedin.com

Autobox User Group - Forecasting & Time Series Analysis, ARIMA, Outliers, Transfer Function & more!
For Forecasting professionals who use Autobox. You can discuss uses of modeling and problem solving and implementation of Automatic Forecasting Software, Autobox.