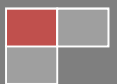2017

Autobox 6.0
User's Guide for Batch Interactive Version

How to Get Data in, How to run, and How to Understand the Output

For use by only the Smartest Forecasters

Preface

About AFS Inc.
- AFS has been in business since 1976 and has many business and academic customers.
- AFS was given the award in the textbook "Principles of Forecasting" as the 'BEST DEDICATED FORECASTING PACKAGE".
- Autobox stands alone at the top of Automated Forecasting Software in the "Daily Data" 2008 International Society of Forecasters Forecasting Competition (**NN5**)
- We are dedicated to giving you a better class of service than other forecasting providers.

Technical support
Customers may contact technical Support for assistance in using Autobox, or even for installation help.
To reach Technical Support, see the AFS Inc. web site at *http://www.autobox.com*

Customer Service
If you have any questions concerning your shipment or account, contact your local office, listed on the Web site at *http://www.autobox.com*. Please have your serial number ready for identification.

Training Seminars
AFS is willing to provide training seminars, through monitor sharing, skype, or even onsite training. All seminars feature hands-on workshops. Seminars will be offered at an as-needed basis, so make sure to contact us if you need help.

Day 1

Introduction to Business Forecasting, Introduction to Time Series, Simple Averages, Moving Averages and Exponential Smoothing

Regression Models for Forecasting, Forecasting Accuracy, Putting it all Together – The Forecasting Process

Day 2

The Model Build Process, Daily Data modeled, Statistical Transformation, Verification of Assumptions, Model Augmentation (ie Stepup/Stepdown)

Day 3

Examples using Customer Data, Complete Walk through of Modeling Process, Verification of Model with Data, Examples using Autobox Data (if time permits)

Table of Contents

Part One: User's Guide

Table of Contents

Part Two: The Autobox™ Reference Guide

Begins on page 108

Introduction

Thank you for purchasing our software. Call us at 215-675-0652, or Skype tomdireill, or email us at sales@autobox.com, if you have any questions.

Autobox can be used with any "time series" data in 5 broad ways:
- **Data Cleansing** (i.e. Identify outliers and correct the data for its errors)
- **Modeling past behavior** (i.e. Did the Promotion Coupon work?)
- **Forecasting** (i.e. Extrapolate a series of numbers into the future)
- **Exception Reporting** (i.e. Which series are out of control? What time period has the most outliers across my different SKUs?)
- **Simulation/Scenario Analysis** (i.e. What would happen if I lowered the price down to $xx?)

If you only have the historical data and no causals, you can execute Autobox either directly from an Excel file**(data formatted as numeric please)** with forecasts written to a file called aboxb.xls or if you have causals(also works without causals) then you can use the "Project Option".

The "Run Excel" option will have Autobox read directly from an Excel file using the series name and data to generate a new table of data. For the "Project Option", the series to be forecasted (and any causal variables) are created in one flat file for each problem set, which will be explained later on in this text. An alternate way to generate the ASC files is to use our inputmacro.zip that comes with the install which will take the data from Excel and create the ASC file for you. The user then runs Autobox by clicking on "Run Excel" or "Batch/Run Project" and chooses how he wants to run the system (using our conditions or making some modifications for your specific purposes. After running the Project Option, you can concatenate the Forecast files (again flat files) that are created which you can then be used for your next step. Again, the "Run Excel" options posts the forecasts to a file named "aboxb.xls".

When using "Run Excel", there is a "Properties" tab that will define the seasonality of the data, how many forecasts and some other info to tell Autobox about your needs.

You might consider this strategy: 1)If you have data with different frequencies (i.e. monthly and quarterly), you should group the series that are alike into DIFFERENT Autobox Project files or Excel files. This is due to the fact that **sometimes** different modeling approaches are used for different data. 2)You also can't mix Univariate and

Multivariate problems into the same Project file so you would need to create two different Projects and run them separately.

Autobox uses automatic modeling heuristics (not pick best) with intervention detection. It tailors the forecast model to the problem at hand including selecting the best lead and lag structures for each input series. It corrects for omitted variables (e.g., holidays or price changes that have affected the historical data that the system has no knowledge of) by identifying pulses, seasonal pulses, level shifts and local time trends, and then adding the needed structure through surrogate variables.

Let's define what a Causal Variable means first. It is BOTH of the following two lines:
Causal Variables – GNP, Price, Promotion, etc.
Dummy Variables – Christmas, Trend, Month of the year effect

Autobox will **automatically** aid the modeling process for weekly, daily, hourly and semi-hourly data. If you have weekly, daily or hourly data, Autobox will add 51 dummy variables for the different weeks of the year. You need at least 1 1/2 year of historical data for this to happen. If you have daily, hourly or semi-hourly data Autobox will add 6 dummy variables for day of the week. If you have hourly data Autobox will add 23 dummy variables for hour of the day. If you have semi-hourly data Autobox will add 47 dummy variables for each half-hour of the day.

For daily data that covers all 7 days(Monday to Sunday), Autobox will different modeling approaches. You trigger Autobox to do this by providing a series name like this "__040106Y11". To tell Autobox to look for these daily effects, just add two "_" before the date and the name of the series where 040106 represents April 1, 2006 and the series(SKU) name is "Y11", for example.

- If a holiday lands on a weekend, Autobox will look for a "Friday before" and "Monday after" effect automatically.
- Search for a day of the month effect.
- Search for an "End of the Month" effect when the month ends on a Friday, Saturday and Sunday.
- Automatically add in U.S. holidays. Note that you can always create variables like this yourself and add them in as a causal variable.
- Monthly Fixed Effects for daily data given that there 78(52*1.5) weeks of data

Autobox will look for Weeks of the Year when the series name does not have an underscore and date "__010108" and 78 weeks of data.

If you have a time series that is not annual, quarterly, daily or monthly, then Autobox will search for **<u>interactions</u>** between "fixed effects" automatically.

If you have an unusual time series in that it doesn't have standard seasonality (ie monthly, quarterly, hourly) you can still model it using Autobox. If you have an hourly data that only covers 20 out of the 24 hours of the day, you can create a special file named '2season.afs" in the installation directory. Inside of this text file place a '20' in it so that Autobox can model it.

Note: If you have data that is very different in scale, we **<u>strongly</u>** recommend scaling your data(by dividing or multiplying) when you have small values and large values. If your Y is 10,000,000 and your causal is .075 then you should scale. You should keep a gap of 6 digits or smaller between the size of the variables (ie 1,000 in sales and causal variable .07 is ok). This is not a "quirk" of Autobox, but rather a common issue for everyone trying to estimate.

AFS has been in business since 1975 and has many business and academic customers. AFS was also selected in the textbook "Principles of Forecasting" as the 'BEST DEDICATED FORECASTING PACKAGE' See page 671 on a your favorite book web site.

Contact us for any questions:

AFS Inc.
P.O. Box 563
Hatboro PA 19040
sales@autobox.com
Phone 215-675-0652
Fax 215-672-2534

The Interactive Batch Version Process: (Univariate and Multivariate)

Interactive Batch Version – Imports XLS files or reads in Autobox "ASC" files and writes forecasts to a flat file.

1) **Import** – Import your data into Autobox from Excel (XLS) or create an Autobox data file (ASC file) – one for each time series in a specific format that Autobox batch requires (more on that later) –
2) **Create an Autobox Project (XML file)** - This defines the location of the data, the series to run and, the output to be stored and where to store the XML file
3) **Add Intelligence** - Create Dummy variables based on your knowledge of historical events + holiday variables + special variables (trading days, monthly effects, etc.)
4) **Choose the series** – decide which series will use the changes made in the previous step
5) **Choose Batch Run** – Runs Autobox
6) **Forecasts Output is generated** – Forecasts, Graphs, and detailed reports on the modeling process
7) **Save?** - Choose whether to save the asc files you have adjusted in step 5.

Process Flow for Running against XLS Files – Univariate only)

Reads in an XLS file and appends forecasts to the same XLS file.

1)**Create XLS data file** – Series are in rows with the SKU in column 1 and the most recent data on the far right see Heinekin.xls(review the Properties sheet and make some adjustments to fit your needs) as a template you should copy in the Autobox directory
2) **Select Graphs and Seasonality** - Choose which graphs you want (ie "Actual/fir/Forecasts) and the seasonality of the data (ie 12 for monthly)
3) **Select a file** - Choose "File/Run From Excel" and select the XLS file to run "HEINEKIN.XLS" for example.
4) **Forecasts are created** - Forecasts are appended into the XLS file. There are JPGs of different graphs and HTM files with reports from the process written to the directory that you might want to review

Limitations:
1) Only uses the history (no causals, holidays, etc.) If you are ok with this limitation which might be ok for your situation then use this option.
 2)This isn't much of a limitation, but we thought just to mention it. Excel has a maximum of 230(16,384 for XLSX) columns so that limits the size of the problem to the number of historical observations plus the number of forecasts. There are 65,536(1,048,576 for XLSX) rows in Excel which is also a limitation.

We will not discuss this option any more in the manual as it is a basic way to run Autobox, but doesn't include causals just yet.

Installation

A directory under the Autobox directory is created during installation named "OUTPUT". This is where the batch will save all of the reports from the run as a default. We include many data files for you to play with along with some XLS files to import as examples. We suggest you open our example XLS files in Excel and take a quick peek to see what the data looks like before importing. ***Click on aboxb.exe to start your Batch session for Interactive or click on go.bat(after editing it for your data first) to run the command version.***

Chapter 2:
Beginning the Interactive Batch Process
Step 1: Getting Your Data Ready

You can either run your data directly from Excel, or import your data from Excel or create an Autobox "ASC" DOS text file. Which way you go is based on what data you have to use. An ASC file is a flat file that lists some specifics for Autobox like (# of forecasts, starting year, starting period, frequency of the data, data type, etc.)

If you have 1,000 series in Excel to forecast and no causal variables then importing using Excel will be your best option so using the "command version" makes sense.

If you have 1,000 series to forecast and causal variables that are the same for every series (GNP, Interest rates, etc.) then importing using Excel will be your best option. Autobox will take your 1,000 series and create 1,000 ASC files. You can then use the "Add Generated Series/Add a User Defined Series" process to bring in your causals to all of the series and Autobox will adjust all of your ASC files automatically.

If you have 1,000 series to forecast and different causal variables that are different for each of the 1,000 series (Price for a given SKU, etc.) then creating an Autobox ASC file will be your best choice. This will require formatting your data using an Excel Macro, Access, Oracle, or SAS to get the data ready for Autobox. See Page 14 "Creating your own ASC file".

There is an example XLS file in the installation directory named AFS20.XLS which you can review and try importing. Your data should be in columns. There are two ways for you to let Autobox use your data. You can either import the data into Autobox using XLS or create your own Autobox ASC file. An XLS file that is included in the installation (AFS20.XLS) for you to review. Note that the last 4 rows are future values.

If you don't have any causal variables then think harder. Is there anything that drives the business that you could include in the modeling? After you identify causals, include the future values of your causals. Autobox can project your causals in the future which is a unique feature in the marketplace, but if you have knowledge of the future you should include that knowledge into the future estimates! Also, if you have dummy variables that you are including in the modeling process like a seasonal dummy model, you do know what those dummy variables would be in the future so include their future values as well.

Importing Your Data from XLS

Open the XLS file that is included in the installation (AFS20.XLS). Note that the last 4 rows are future values. It is important to note that the causal variables will have a data type (which defines for Autobox how the future values are created and used in Autobox – more on that on the next page). The important thing to note is that if the series have no future values then the data type is defaulted to '0' which means Autobox will forecast the future values which may be what you want, but depending on your variable may not be what you want. If the causal variable does have future values then the data type is defaulted to '2' which means you supply the future values. If you want a '1' or '3' data type you can't use the XLS import facility and are forced to create your own ASC file as you can't change "en masse" data types. It just isn't possible to manage all of that data as Autobox is not a database. Note that if you have a promotion that if you have daily data and a promotion that goes on for many periods you would want to use a data type of '1' as you can't really look for lead/lag from multiple time periods.

Let's define the data types now so you will understand how important they are:

<u>Data Types:</u>

- 0 – Autobox will forecast the future values of the causal series and try and identify current and lag effects during the analysis

The next 3 have the future values defined by the user AND:

- 1 – current time period effects (use this for something like "seasonal dummies" or you know there is no lag/lead effect) are analyzed

- 2 – current time period and lag effects (same as data type '0' except the user is specifying the future values) are analyzed

- 3 – current time period and lag and lead effects are analyzed

In order to change the data type default from a '2' after importing, you will need to import future values of causal variables (which would logically be at the very bottom of the dataset see the last 4 rows of AFS20.XLS as an example). Choose "Series/Series Information" to modify the data type.

Autobox will read in your XLS file and ask for you to tell it some specifics like when your first data point begins, the frequency and the number of forecasts you are looking for.
Our Excel import is used for importing data from spreadsheets that are vertically oriented. This means that series are in columns and observations are in rows extending down.
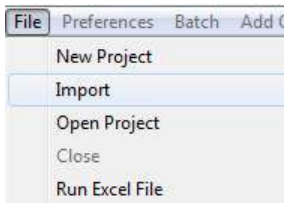<u>***You must have the data beginning in row 1 or 2!***</u>

Let's Choose File/Import.

Fig. 2-0 – Import Excel File

Let's import AFS20.xls(included in the installation) and click "open".
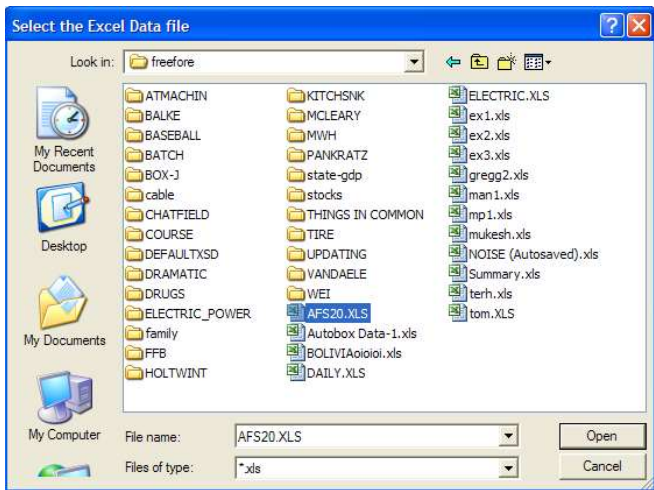

Fig. 2-1 – Selecting AFS20.XLS

We need to tell which sheet to get the data from and our data is in the first sheet so we just click "next".
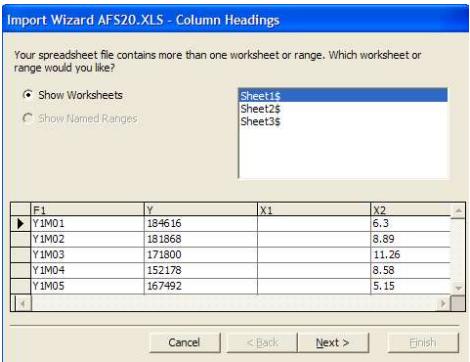

Fig. 2-2 – Importing Data to First Sheet

We have titles in the first row so we leave the box checked and click "next".
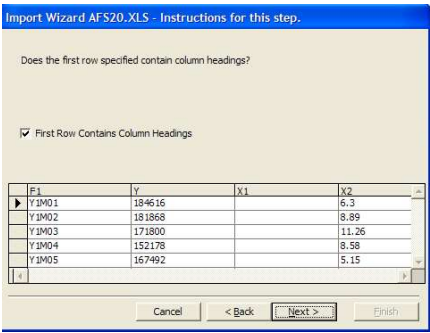

Fig. 2-3 – Importing Data

We will skip the first field as dates are not in the needed and defined in the next step so we will click on the box to remove F1(default name assigned by Autobox as it did not have a tile). Not shown here, but we also choose to remove X1 as only half of the observations were included. You click anywhere on the "X1" column to highlight it and then again click on the "Do not import field (Skip)". You might notice that there is only partial data in X1. There is another way to approach this and that is to remove the older data that did not have X1 values if you felt was so important in describing Y. You can run Autobox both ways to see the differences in approaches. Click "next"
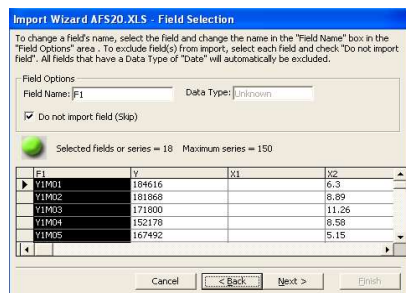

Fig. 2-4 – Dropping Unnecessary Columns

Notice that fields F1 and X1 are now dropped. Click "next".
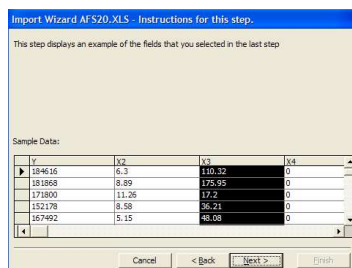

Fig. 2-5 – Data with Removed Columns

You have the option to filter the rows being read (like Excels' filter feature), but we don't have a need for this here. Click "next".
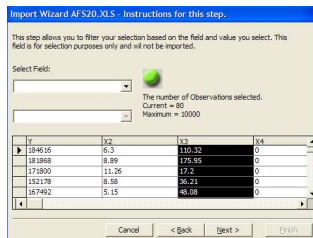
Fig. 2-6 – Continuing Import

Click on the arrow and choose 'Y' as the output variable you are trying to predict. Click "next".

Your Excel spreadsheet may be a causal problem or it just might be many different problems. If your data is not a causal problem and each column is a different problem, just pick any field as the output series field as it won't make a difference. Autobox will save the ASC file(s) into the folder where it imported the XLS file from. Autobox will also automatically generate an ASC file with all columns for causal analysis and also an ASC file for each and every one of the columns in anticipation of the data NOT being a causal dataset.

Fig. 2-7 – Setting the Properties

The following explains the individual properties

Output Series Field – For Causal problems, choose the series to be predicted, otherwise choose any field.

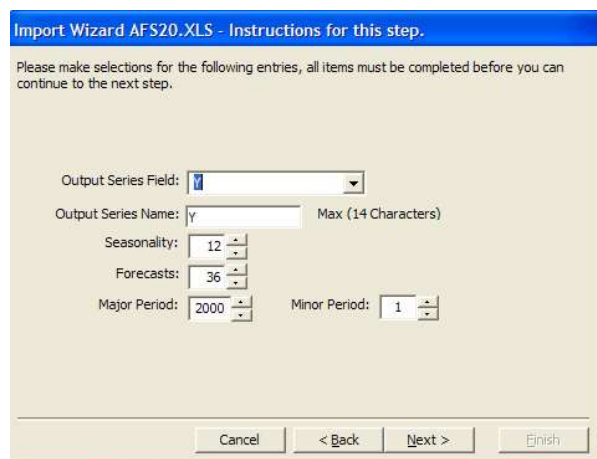**Output Series Name** – This is prefilled when you import from Excel, but you can change it if you want to.

Seasonality – How often the observations are taken. This is a good time to layout the different seasonalities for Autobox – 1 for annual, 4 for quarterly, 12 for monthly, 52 for weekly, 7 for daily(7 days in a week), 5 for daily(5 days in a week) and 24 for hourly.

Forecasts – The number of forecasts you would like Autobox to generate

Major Period/Minor Period - Let's give a few examples of Major/Minor Period to help you see – For February 1990 monthly data, choose 1990 for the major period and 2 for the minor period. For January 2, 2006 daily data, choose 1 for the major period and 2 for the minor period.

Known issues and limitations:

       • The spreadsheet must be closed in order for the import to work correctly. This means it can't be open in Excel while you are trying to import.

       • Data must be in columns.

       • Series names are limited to 14 characters.

       • Vista XLSX files will not import.

Here is an audit of what was read in. Note that the last 4 observations are future values of the causals to help with forecasting. Click "finish".
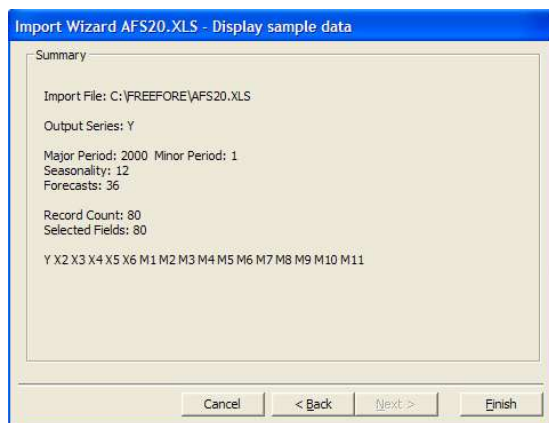

Fig. 2-8 –Imported

We saved the file as "itisreadin.asc". Click "Save" and a message informs you that it was saved.
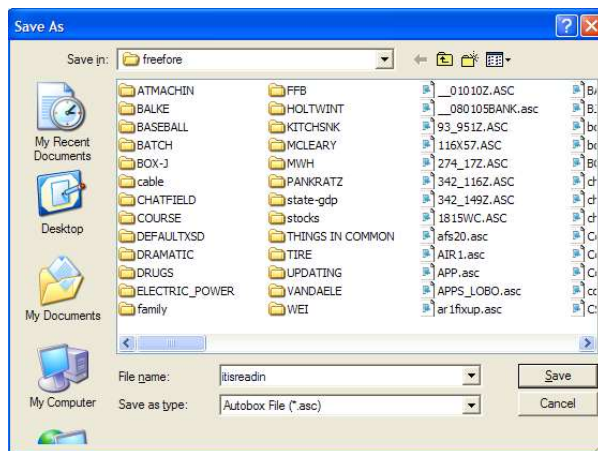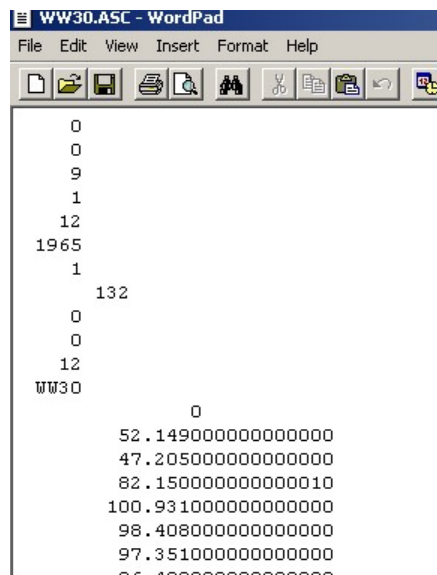

Fig. 2-9 – Save the File

Create your own ASC file (Brief Version)

We include datasets to show you how to prepare data for Autobox. The files must have a name of "ASC". Here is an example of "WW30.ASC" file (which you should use as a reference). Information like the # of series, the length of the series, # of forecasts, starting month, starting year are defined here. We have a full explanation in the **_appendix_** on how to create an ASC file. This data preparation takes time so it would make sense to get that process underway!!!

Fig. 2-10 - .ASC File

You will need to convert all of your data from SAS, Oracle, Excel, etc into an "ASCII" or "Flat file" format. Each dependent time series will need to be saved into a separate file for each series to be modeled. Note that if you have supporting series e.g. price, promotions, events these also need to be stored in each of the files. The file needs to have an extension of ASC (i.e. ww30.asc). There are some header records needed at the top of the series to define for Autobox batch details of the series (i.e. name, frequency, etc.). Future values of casuals are put at the very bottom of the file.

Here's a trick when you have long lagged response to a marketing event or "Adstock" type of data. For example, drilling done at an Oil Well will also have long lagged responses to injection of water and chemicals amongst similar type of data. If you specify the causal series name with the words "DYNA14" and place a "1" when the event occurred and leave all of the other data as a "0", Autobox will react by searching for a decay impact over the next 14 periods. Remember that the version of Autobox you are running might be constrained to 6/30/150 variables and you might blow out the array of the program if you are careless here. Note: You can only have names as long as 6 characters where the first 3 are "DYN" and the fourth is a letter or a number and 5th and 6th is the lag structure.

 You will also need to put a unique character for each DYN variable. Note that we put an "A" after DYN to make the name unique.   The max lag is 50. If you have a lag of 4 you need to put 04(ie DYNG04)

 Note that the "data type" MUST be '1' for this to work.
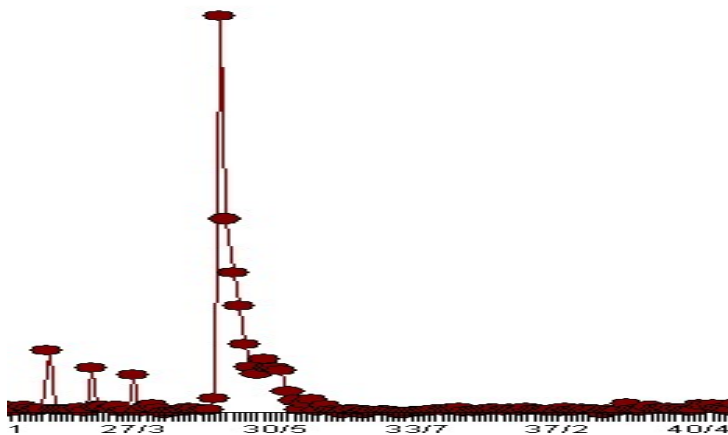
Fig. 5-0 – Example of a Long lagged impact

Step 2 - Create an Autobox Project

You will need to define where the data is located and the series to be analyzed and where to store the results. This is called an Autobox project file and has an ".XML" file name.

You choose "File/New" to begin.
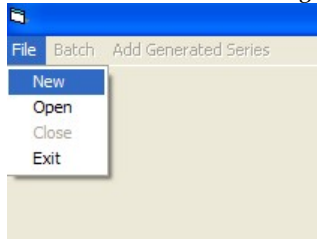
Fig. 2-11 – New Project

We need to tell Autobox which directory to find the data and we point to all of the Box Jenkins time series from the classic text book. Click 'next'.
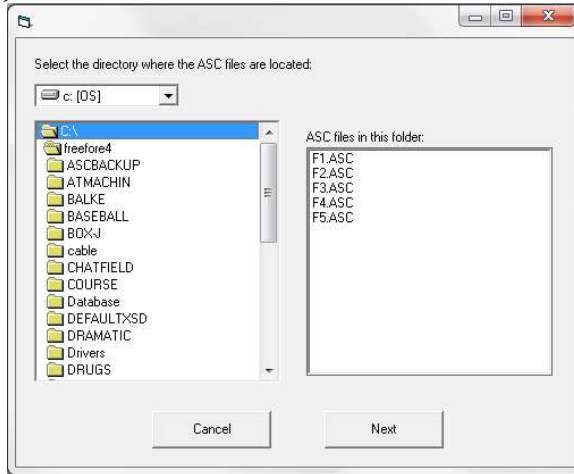
Fig. 2-12 – File Selection

We only wanted to run the first 4 series so we highlighted them and clicked 'add' which moved them to the right and clicked 'next'.
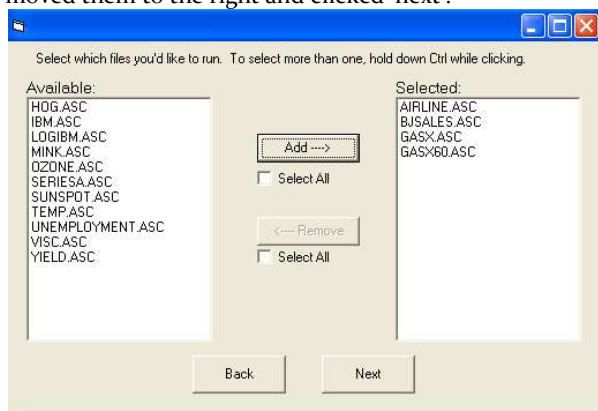
Fig. 2-13 – Adding Series

We recommend you save the work files Autobox creates(which aren't really need by you) to a working directory that was created during installation named "work_area". Click "save". Note the directory OUTPUT under "work area". This is where the output will be saved.
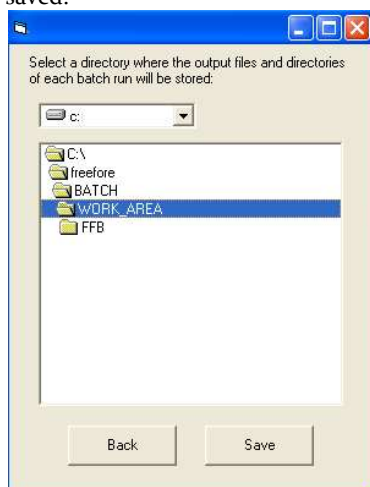
Fig. 2-14 – Select Output

You should keep all of your Autobox projects in one spot for organization sake in the batch directory. Click 'save' and a "file successfully message" is generated (not shown below). Click "ok"

Fig. 2-15 – Saving the Project

Now we run Autobox by choosing "batch/run".
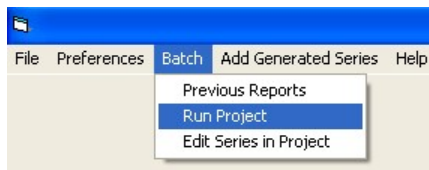
Fig. 2-16 – Batch/Run Project

And we have results. The results and reports are written to a flat file on your computer as explained at the beginning of the manual. The forecasts are all in one file (stored in the directory that contains the output) that has the name of the XML file but with a "PRO" file name. So, if the name of the XML file was "example.xml" the file would be named "example.pro".. Various reports and graphs are also created for your review.

As for the reports: Details.htm has an audit of the modeling steps. Adjust.txt shows the historical cleansed of outliers. Equation.txt shows the final equation used to forecast. Intrvent.htm reports that intervention and the impact. Stat.htm reports the statistics for the model fit and t tests for the variables.

The graphs are self-evident.
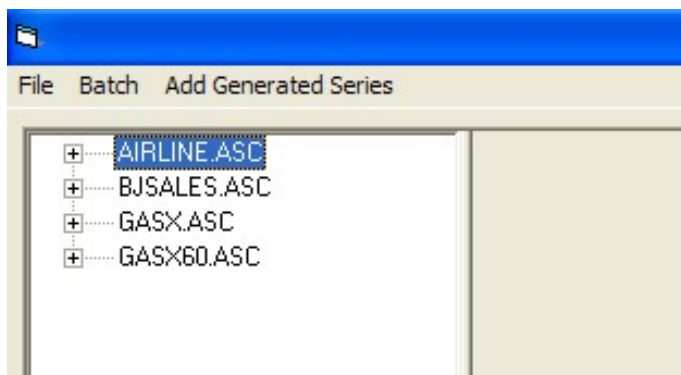
Fig. 2-17 – Selecting Output to Review

This is the graph for the Airlines series.
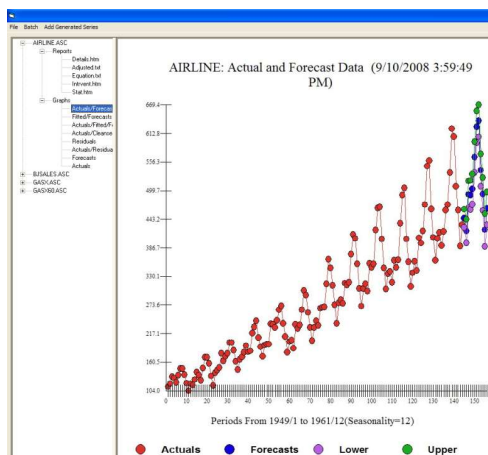
Fig. 2-18 – Generated Graph

Step 3-Add Causal Variables

You can use Autobox to add important causal variables to be included in the modeling process. Choose "Add Generated Series" and you will be asked to select a series so that Autobox can find out what the properties of the series are and to be specific the seasonality (i.e. monthly) of the series. Once this is established then Autobox will offer causal variables that are appropriate to your problem at hand. It does not mean that any of these variables will be included in the model just that they will be considered to be in the model. It can be very powerful if you have "apriori" information like known interventions in the history or even in the future and include it to aid the model in the process. **Note that there is a directory titled "ASCBACKUP" that contains the ASC files before you added the causal variable in case you dislike the results WITH the causal.**

You should note that for level, seasonal pulse and local time trends you can either allow the event to occur into the future or you can define the point that it has ended.

Here we select a daily series.
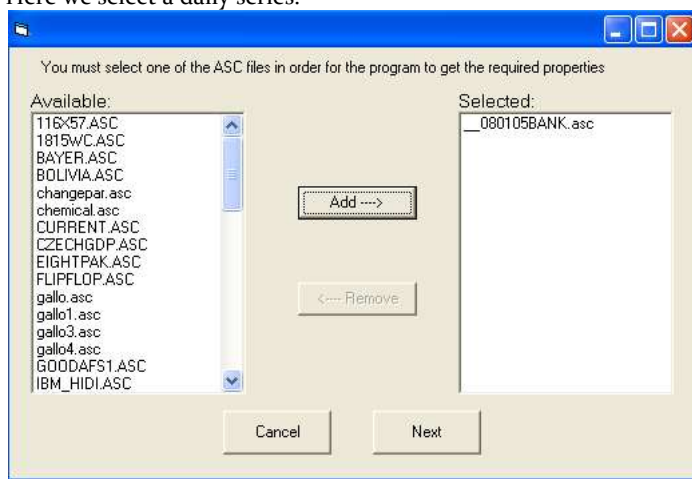
Fig. 2-19 – Adding Causal Variables

If you had annual data then you would not be allowed to add a seasonal pulse variable into the analysis as it makes no analytic sense to do so as the data has no seasonal fingerprint to it. Here we have all options selected for **__illustrative__** purposes only. Click "next".
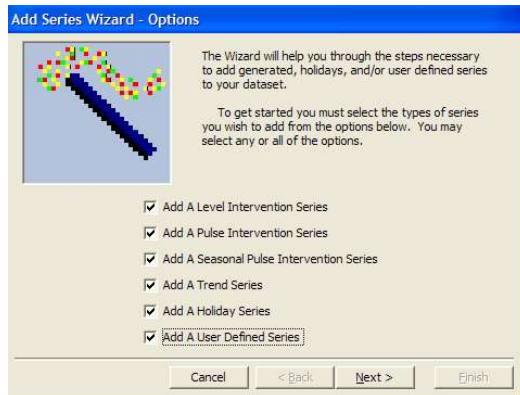
Fig. 2-20 – Add Series Wizard

A level intervention is a sudden increase/decrease in level of the mean that continues into the future. A dummy variable is included in the dataset that looks like this 0,0,0,0,0,0,0,1,1,1,1,1,1,1.

Now, since we are running in batch mode we are suggesting that every series in the batch run had a level shift at the 70th period. This might be the case as a competitor may have gone bankrupt causing your sales to have a one time increase in the mean which would apply to all of your product lines. On the flipside, you might have an outlier for just one series so you might want to move that series into its own project and run it by itself with the level shift variable. If you do not and bring a level shift variable for series that didn't experience a level shift you might change the model created so we recommend that you don't to this. This applies to all interventions (level shifts, pulses, seasonal pulses, local time trends). You can test each way out for yourself and give us feedback from your experience for our curiosity.

Here we are telling Autobox that at the 3rd day (in the period box) in the 70th week (in the year box) had a level shift (note that the seasonality is 7) and that it continued to the end of the dataset as we did not adjust the ending dates which default to the very last observation. Let's assume we have a shirt factory and we added an embedded electric light in our collars and people like it and end rushing to the store and buying our shirts and continue to show up each day to buy these shirts.

Fig. 2-21 – Setting Interventions

A pulse intervention is a one-time sudden increase/decrease which occurs at only one point in time. A dummy variable is included in the dataset that looks something like this 0,0,0,0,0,0,0,1,0,0,0,0,0,0,0.

Here we create a pulse variable at the 45th week and 4th day. Here the factory lost power and no shirts were made at the factory and everyone had a siesta.

Fig. 2-22 – Setting a Pulse Variable

A seasonal pulse intervention is a cyclical increase/decrease which occurs at only one point in time and reoccurs. A dummy variable is included in the dataset that looks something like this 0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0.

Here we create a seasonal pulse variable at the 63th week and 2nd day. Sales are now low on Tuesdays because the local church decided to start a bingo game in the town next door which is hurting our sales.

Fig. 2-23 – Seasonal Pulse Variable

A trend intervention is a sudden change either increasing or decreasing in the slope which occurs at one point in time and into the future. A dummy variable is included in the dataset that looks something like this 0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,2,3,4,5,6,7,8,9

Here we create a trend variable at the 89th week and 5th day. Sales are now increasing on steady stream downwards as war has broken out in our county and people are afraid to come out of their houses to buy shirts.

Fig. 2-24 – Trend Variable

Holidays often have big impacts on many lines of business so we recommend including them to see what effects they may have on our business.

Our example is daily data and we have selected all 16 holiday variables. FYI: For monthly data, you only have the option of selecting 12 holidays as you are limited to their being only 12 months in a year.

Fig. 2-25 – Holiday Variables

Last but not least is the "User Defined" variable which depending on the seasonality of your data you may have different options of including different variables. For example, while we have daily data there may be some quarterly variation that corresponds to the season due to weather that we would want to account for in the modeling process. So, we would select the quarterly variable.

We must **<u>caution</u>** you that there are two options in this example that do not make sense to choose for daily data. There are two on this list you would not want to select as logic would have it. Day and Business Days make no sense to use for daily data. Business days is really only meant for monthly data. A day of the week effect is already included when Autobox analyzes daily data(go back and read page 3 if you are lost here!) so there is no need to include another day variable. Autobox would drop yours if you brought in a duplicate variable anyway as a self-protection scheme.

Fig. 2 – 26 – Add User Defined Series

So, we are going to include a quarterly variable and Autobox fills in the name automatically. We now must pick a data type. If you think that there is a "lead effect" to this promotion where demand is shifted before the event then this would suggest that the type should be set to '3', but it would be not be logical to have three dummy 1's consecutively when you are looking for a lead and lag relationship. There should just be 1 dummy set in March and the February and April should be a '0'. Autobox looks for lead effects up to 6 periods before the '1' dummy is specified. As a protective measure, Autobox will change your type from a '3' to a '1' if you have 1's that are within 6 periods of one another so be careful to follow these previous words or call us to discuss.
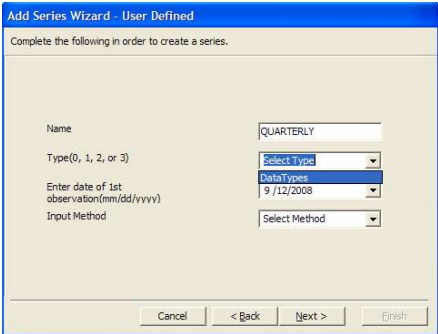
Fig. 2-27 – Setting Data Type

As discussed earlier, it would not make sense to choose 2 or 3 as there are variables generated for each quarter so there is no need to analyze if a lead or lag effect is necessary. Option 0 is not logical as you would not want to forecast a dummy variable. Choose 1 and click "back" to continue (strange as it sounds to click "back", but just do it).
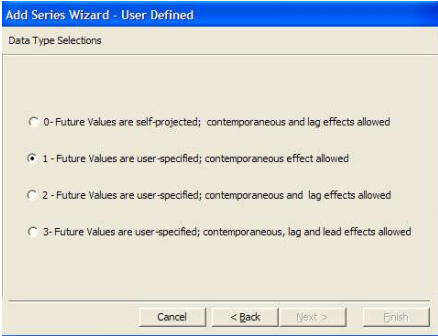
Fig. 2-28 – Data Type

We need to tell Autobox the actual date when the data starts so that the dummy variables can be created properly using the calendar. We click on 4/10/2007
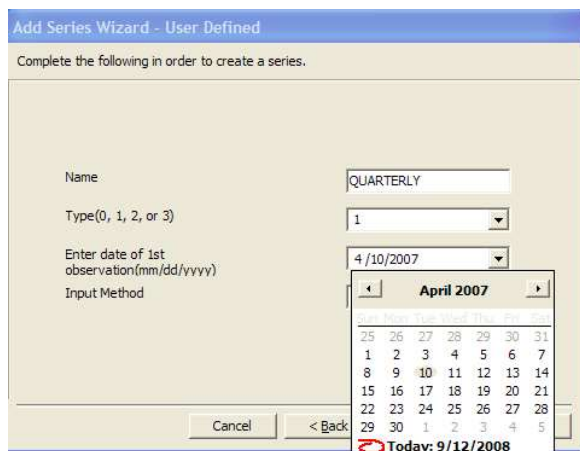
Fig. 2-29 - Calendar

We click "next" and then "finish" (not shown).

After this, we are asked which series we want to use this new causal variable with and we select all the series (assume they are quarterly in this example) and click "next".
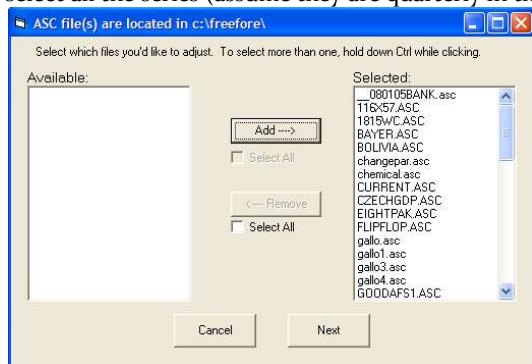
Fig. 2-30 – Add to Series

Now, if you didn't read page 3 and you have a series with different seasonality and you just plowed ahead and tried to go through this process without grouping your series into different projects by seasonality you will get this error message. You can still go on to run Autobox, but the reality is that you are not optimally modeling some of the series in this potpourri of different series and should create different project files (XML) for each type of seasonality.
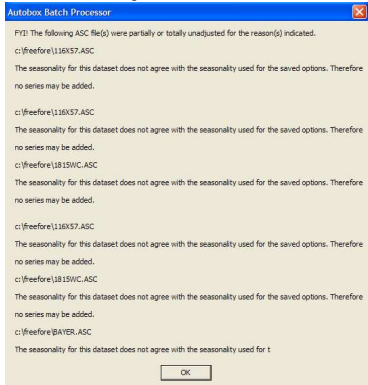
Fig. 2-31 – Error Message

Now we run again through Batch/Run Project (you can cancel the batch run by clicking on 'cancel').

You can see the progress as each series is completed, but you can't review any of the series modeled until after the batch run is completed.
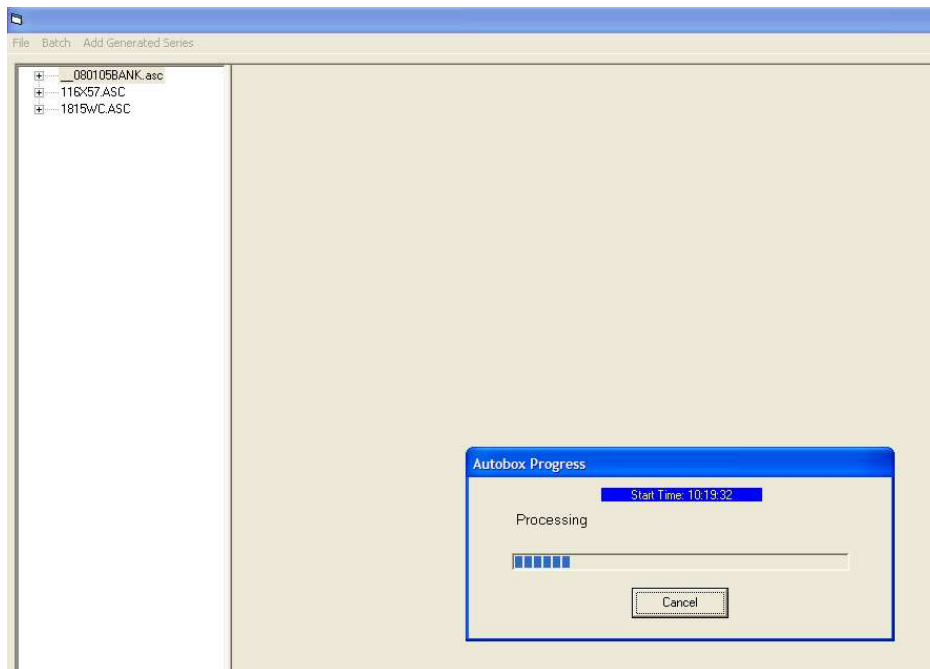
Fig. 2-32 – Progress of Autobox on Series

We added casual variables to the process and Autobox asks for confirmation that you would like the ASC files to be adjusted to include these variables. Select the series and click "add" and then "next".   A message will inform you that ASC files were modified.
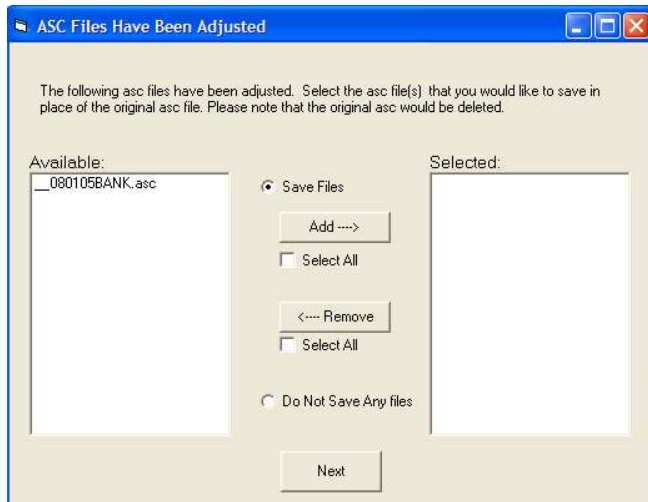
Fig. 2-33 – Files have been Adjusted

Let's talk about how to build your own dummy variable a little bit more. If you had a promotion starting in February and it ended in April, you would want to put a '1' for those three months to indicate the event.

Period/Time	3	M_PROMO
1 2008/1	4186.00	0.00
2 2008/2	3462.00	0.00
3 2008/3	2552.00	0.00
4 2008/4	3045.00	0.00
5 2008/5	1287.00	0.00
6 2008/6	1395.00	0.00
7 2008/7	1290.00	0.00
8 2008/8	1681.00	0.00
9 2008/9	1435.00	0.00
10 2008/10	931.00	0.00
11 2008/11	401.00	0.00
12 2008/12	588.00	0.00
13 2009/1	2225.00	0.00
14 2009/2	15399.00	1.00
15 2009/3	9840.00	1.00
16 2009/4	22822.00	1.00
17 2009/5	6356.00	0.00
18 2009/6	3070.00	0.00
19 2009/7	2709.00	0.00

Fig. 2-34 – Generated Data

There is one trigger based on the name of the series: There is one very special Autobox trick that we want to discuss up front. For example, if you have a "dynamic" promotion over a period of two weeks (and you had daily data) that causes demand to shoot way up and it slowly ramps down back to the mean (decays). If you specify the causal series name with the words "DYN14" for example, Autobox will react by modeling the promotion to decay over the next 14 periods the promotion was running. Note that the data type(SEE THE SECTION ON CREATING YOUR OWN ASC FILE FOR MORE ON DATA TYPE) MUST be '3' for this to work. Also, if you have a "patch of outliers" that are in the same range (ie all zeroes) then you can use a '1' indicator during that patch. If you have a "patch of outliers" that varies wildly (ie high, low, etc.) then use "DYN" and the length and again data type must be equal to '3'.
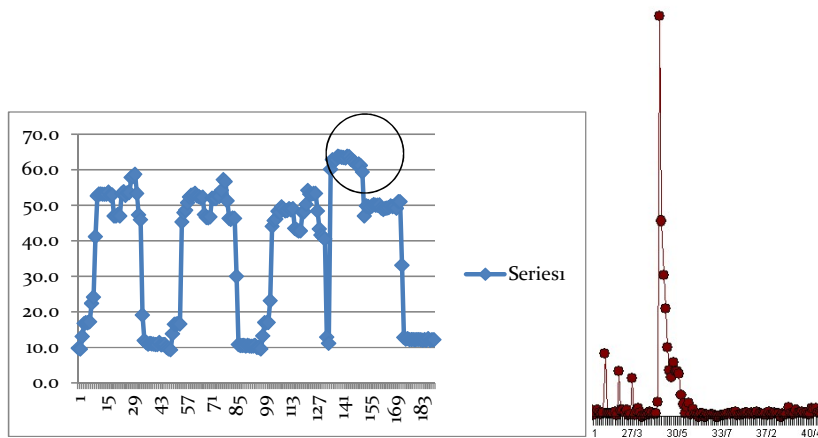
Fig. 2-35 – Graph showing outliers

Step 4: Executing using an Excel file

When you run directly from an Excel file your Excel file data should be in rows. The first column ('A') should have the name of the series and the data in columns 'B' an d on. The oldest data should be in column 'B' with the most recent data in the far most right column. We deliver an example with the installation named Heinekin.xls that shows how the data should look.

Choose File/Run Excel File and tell which XLS file to run.
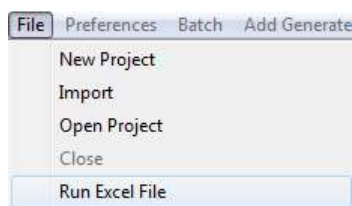
Fig. 2-36 – Run Command

You can then select which graphs you want to see. We like the actual, fit and forecast graph.
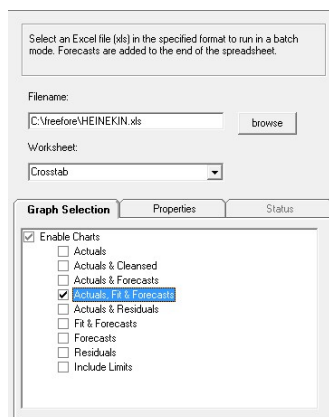
Fig. 2-37 – Graph Types

You can also select the seasonality, beginning period information and forecasts. If you have weekly data specify 52, annual 1, monthly 12, quarterly 4, daily with weekends 7 ,daily with no weekends 5. The beginning period isn't import to the analytics and it is only used for labeling the reports.

Click on "run" and Autobox will it will write out a file called aboxb.xls with the forecasts appended to the right of the spreadsheet.

Fig. 2-38 – Setting the Properties

Step 5: Output

When Autobox has finished running, it automatically brings up the reports and graphs generated during the processing and tells you where this data is saved to disk.(They are stored in a "working" directory named "OUTPUT" which is automatically created wherever you have saved your XML file). You will need to delete the files in "OUTPUT" once in a while as they will just keep accumulating!!!! A file with all of the forecasts for the batch run is created and saved to disk with a "PRO" file name. The combined forecast file is in the "OUTPUT" directory and has the same filename as the xml file. So, if the xml file is "test.xml" then the forecast file will be "test.pro". Now if you ran 1,000 series and you then go to "edit series in project" and choose 5 SKUs and you wanted to try a different methodology on and you rerun. NOTE: The old 1,000 reports will not be shown in Autobox, but they will still exist in the "OUTPUT" directory.

Reports

Detail.htm – Tracks the steps and decisions to create the model

Intrvent.htm – Lists all of the detailed information on interventions

Rhside.txt – Lists how each of the variables numerically contribute to the forecast.

Stat.htm – Shows the statistical fitting statistics (RMSE, AIC, etc.) and model with P-values

*.pro – The forecasts for all series

Verbal.txt – The model explained in "English"

Pulserpt.csv, Levelrpt.csv, Trenrpt.csv, Parchrpt.csv. Varchrpt.csv, Critiqu1.csv and Critque2.csv. – see the next few pages for more discussion on these!
filename.ADJ - Historical Data cleansed for outliers (found in installation directory)
filename.SAF - Safety stock calculation for 90% service level (found in installation directory (put nosim.afs in the autobox folder to get this))
Doparcon.txt – Reports the Parameter Constancy test (ie Chow test).
Modified.123 – Series cleansed of outliers

*.HIS – The actuals and forecasts ready for import into a tool like Tableau/Spotfire/Qliqview to graph and do further analysis or "BI"

*.PMM – When a file called "weights.afs" exists in the directory, a 4.1 PMML file is generated with the model built by Autobox ready to be imported into a post processor. See more at www.dmg.org Copy all pmm files to pmml after a run using the DOS copy command (ie copy *.pmm *.pmml)

Formatted: Indent: Left: 0"

*.TRA – Ready to be imported to BI Tools like Tableau, Spotfire, etc. with name of series, actual, fit, forecasts, confidence limits, flag indicating if seasonality was found. These files can be concatenated easily using the Dos copy command into one file to be imported.

Click on the windows Start Icon (bottom left of screen) and type 'cmd'. Create a file with one blank line named "dummy" using notepad or text pad then issue the next command to concatenate all TRA files into one file::
Copy dummy+*.tra all.tra

*.OUT – Reports all of the outliers sorted by size.

*.PEK – Shows the largest and and smallest values in the forecast. It's called "Peaks and Valleys" to show what the highs and the lows might be in the future.

*.PMD – When a file called "weights.afs" exists in the directory, a 4.1 PMML file is generated that is the dataset accompanying the pmml file ready to be imported into a post processor. See more at www.dmg.org

Graphs
JPGS - Actuals/Forecast, Fitted/Forecast, Actuals/Fit/Forecast, Actuals/Cleansed, Residuals, Actuals/Residuals, Forecasts, Actuals, ACF Originals,ACF Residuals
Tip:QC Autobox's output by reviewing the residuals graphs and the graphs of the actual fit and forecast. We use Google's Picasa to go through a slideshow(you can also use Windows Slideshow). Just copy "*actfitfore.jpg" and copy those files to the "My Pictures" folder and use Windows to do a slide show to QC the graphs.

Step 6: Exception Reports

Autobox creates a number of reports to help you understand your data and track the quality of the forecasts being generated so that you can review them for accuracy.

- Outlier Exception Reports - Macro view of where outliers occurred which can suggest that "maybe something happened in the history". It may also trigger you to realize that it was due to a marketing campaign and then realize to bring these in as causal variables when modeling the data into the future
- Forecast Monitoring Report - Quick check to make sure the Autobox forecast is reasonable vs. a simple averaging method of your choosing (ie 4 period equally weighted average) as a baseline for comparison
- Forecasting Fit Exception Report - Compare Autobox fit vs 2 simple methods

Chapter 3:

Accuracy, Models, Errors and ASC Files
Outlier Exception Reports

Pulserpt.csv, Trendrpt.csv, Levelrpt.csv, Parchrpt.csv, Varchrpt.csv – Log files that show a Table of a pulse(or trend or level or parameter change or Variance change) outliers at different time periods(see the first 11 rows to see what it looks like in the picture below). If you have 200 series and you find that 150 have an outlier at time period 02 then it might cause you to think about what happened at this point in time that you failed to include as a variable in the model from the beginning for these 150 series(possibly for all 200 series?). In a couple of steps you can find if these occurrences also occurred annually suggesting that it was a holiday that was omitted in the modeling process. Open the file in Excel and sum each of the periods. Copy and transpose that row to a column. Create the counting numbers next to this column (1,2,3,4, etc). Sort the two columns by the count largest to smallest. Now you have the count of the time period with the most outliers at the top. Below is an example with 10 SKU's with 1,049 daily observations. We did some investigation by subtracting different time periods to identify a missing holiday variable but we didn't find any differentials of 365 so given that we conclude that these

are just interventions and not a systematic pattern since 3 out of 10 could randomly occur at a given time period by chance. Note that the series need to all start at the same time period so that the data is aligned! Also, this file is a log file so you need to delete it once in a while to keep the size down.

The level and trend don't show the sign of the coefficient making their report more ambiguous. You can turn off these reports by creating a file named "nocsv.afs" in the installation directory. This file is a log file so you can delete it once in a while to start fresh.
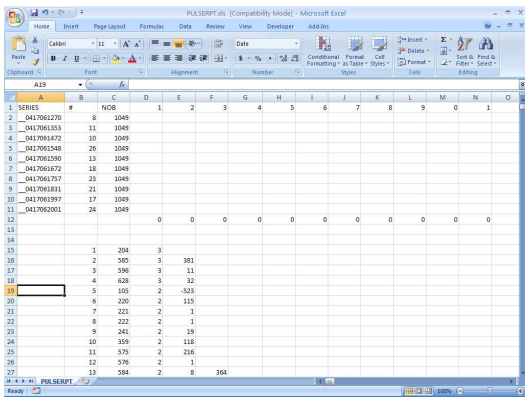
Fig. 3-1 –Exception Report

Early Warning System Report

The report "earlysig.txt" is created to help find out if the last observation is "out of control". The report tells you the name of the series, the last observations number, the probability of out being out of control, the observation, and what the observation was expected to be. There is one record added to this file every time (up to 20 series and then the file is purged to avoid a large file—the batch version will continue to write out to this file so if you have 50,000 series this file will have 50,000 records). Note: This report is saved to your installation directory NOT the OUTPUT directory.

You can bring this file into Excel and sort on probability (ascending) to find the series that seem to be "out of control". I ran the series inlier and there was nothing found to be "out of control in the last observation" as you can see below. To investigate further, I changed the last observation from a 9 to 5,555 and then reran Autobox. The second row shows a low p-value to show that there is something wrong. It prints out what the value should have been in the expected column. ***Note: This file is a log file so you can delete it once in a while to start fresh.***

```
EARLYSIG.TXT
| ITEM          NOB PROBABILITY        ACTUAL              EXPECTED
inlier                                 9.0000000000
inlier          10  .0000        5555.00000          5.0000000000
```

Fig. 3-2 – Early Warning Report

Forecast Monitoring Report

This report gives you a way to check that the forecast from Autobox compared to a simple method to locate if there are any forecasts that are very low or very high. The report takes the ratio of the two forecasts so that you can locate very low or very high ratios to inspect. This report has many false positives as a simple method is in fact that, simple.

You can create a file named "fore-mon.afs" in the installation directory and if you wanted your simple forecast benchmark to be a weighted average of the last 4 periods you would have a total of 5 rows in the file:

4

.25

.25

.25

.25

You would save this file and run Autobox and a report would be created named "fore-mon.csv". You can sort on the field "Autobox/base" to identify low and high differences in baseline forecasts. This file is a log file so you can delete it once in a while to start fresh.

ITEM	NOB		BASE	AUTOBOX	DIFFERENCE	AUTOBOX/BASE
REGION1	147		1025	500	-525	0.49
REGION10	147		0	341		
REGION11	147		225	58	-167	0.26
REGION12	147		12516	11477	-1039	0.92
REGION13	147		7642	7072	-570	0.93
REGION14	147		2342	4739	2397	2.02
REGION15	40		0	19		
REGION16	142		150	0	-150	0
REGION17	147		7213	6319	-894	0.88
REGION18	131		424	611	187	1.44
REGION19	91		0	0		

Fig. 3-3 – Forecast Monitoring Report

Forecasting Fit Exception Report

This report gives you a comparison of the fitted MAPE between the Autobox model, the naive model and a mean model. You create a file in the installation directory named "mapecomp.afs" and a report will be generated named "mapecomp.csv". The Autobox model column could be sorted and used to compare to the simple methods to provide a check that Autobox is doing better than the simple methods. Of course, there still remains the possibility for false positives. This file is a log file so you can delete it once in a while to start fresh.

ITEM	NOB	AUTOBOX MODEL	NAIVE	MEAN
REGION1	147	0.374	0.911	0.724
REGION10	147	0.225	0.376	0.162
REGION15	40	0.046	0.224	0.228

Fig. 3-4 – Exception Report

Measuring Accuracy

This utility is for evaluating forecasting accuracy from many withheld periods. The number of forecasts in the ASC file will determine how this will all behave.

Create a file in the Autobox folder called "Rolling.afs" using "Notepad". There will be 2 rows in the file. The first row tells the system how many period back you want to withhold. The 2nd row defines the periods out between each withheld amount. If you have 23 as the number of forecasts and have a 1 in each of the two rows, one snapshot will be produced from 23 periods back.

You would typically want the second row to be a "1" so you could evaluate each period getting a "rolling accuracy" from many origins, but if you wanted to get only every other period you could put a "2" for example, but this isn't usual to do. The # of forecasts in the ASC file needs to match the second row. In this example, we used 20 and a 1. If we have 144 observations and we have 12 forecast periods, the first observation you can withhold is 132 as you can see from the file created "summape.csv" as shown below. The max is 52 for row 1 as the maximum seasonality is 52 in Autobox so 52 is the furthest out you can run a forecast when doing this.

20

1

ITEM	NOB	NF	MAPE	ACTUAL	FORECAST
BJ07	132	12	2.75	5714	5797
BJ07	131	12	2.74	5687	5649
BJ07	130	12	3.27	5659	5540
BJ07	129	12	3.34	5605	5526
BJ07	128	12	3.6	5560	5462
BJ07	127	12	3.41	5513	5427
BJ07	126	12	4.73	5439	5218
BJ07	125	12	3.61	5376	5274
BJ07	124	12	3.29	5324	5285
BJ07	123	12	3.73	5259	5101
BJ07	122	12	5.18	5246	4974

Autobox will run many different analysis from different origins and generate a CSV file to show accuracies. You will need to delete this output file (ie summape.csv) after your analysis as it will just grow and grow. The HTM files generated on the disk are the analysis files if you want to review. This can't be used with Simulated forecasting. This can be done using causal problems, but with no searching for day of the month impacts (ie nofixday.afs) due to complications.

Saving and Re-using Models

If you need to reduce the time spent on remodeling when new data comes in, you can tailor the process to NOT remodel and just *reuse* the previous model and use it with the latest data. The **_downside_** is that you are not using the latest information to make decisions. We would caution you to use the latest model, but this decision entirely depends on you.

How can you reuse models in order to cut down the time expense of remodeling? In order to do so, you need to model and save it for later use.

To save your model(s):

- Create a file named <u>savemod.afs</u> in the installation directory
- Run Autobox batch and the models will be created and saved as *.MOD
- Create a zip file of all of the *.mod files for later use

To reuse your model using an existing model three things are necessary:

- We will assume it has been some time since you created the model so you should unzip the MOD files to the directory.
- To tell Autobox batch that there is a model saved and should be used:
    a. Edit Engine.xxx and change line 37 to a '1' and save as Engine.go or
    b. create a trigger file called USEMOD.AFS in the installation directory
- The actual model files (*.mod where * is the series name) must exist
- If you want to override AFS' expert system, edit Engine.xxx and modify any of the conditions as you see proper (see Appendix for more on Engine.afs) and save the file as "Engine.afs". Also, see predict.xxx for an example where we have restricted the analytics to just forecasting (note that estimation is turned off by setting it to 'o'). Also, note that the structure is hierarchial. If an item is set to zero than any item that is indented underneath is also not going to be executed.

Here is how to reuse models for the "Comprehensive and Multiple batch modes"

In order to let Autobox batch know if you would like to save a model you need to place a 0,1,2,3 after the name of the zip file. A 'o' runs the entire process and saves the model only. If we wanted to save our model we could put a 'o' as seen here.

Next, we will issue the command to re-use the model.

Issue the command at the DOS prompt:

OUTPUTATCH1 WEI3 0 (see next section on changing default modeling conditions).

When a '0' is used to run OUTPUTATCH1 a file is saved with the model. The name of the file is the series name with a "MOD" as the extension (i.e..We30.MOD). The MOD files are copied to MOZ files for backup purposes as the MOD files are deleted after being used via the model reuse (type 1,2,3). The contents of the saved model file (MOD) is discussed in Appendix 1 in the section "How to specify your own model".

The three options to put at the end of the line which use an existing model

Option/Goal	File Used
1. SIMPLY FORECAST	PREDICT.XXX
2. TUNE COEFFICIENTS AND THEN FORECAST	TUNE.XXX
3. TUNE COEFFICIENTS & POSSIBLY REMODEL THEN FORECAST	REMODEL.XXX

If we run Autobox batch with a '1' to run the R-squared is NOT CALCULATED as estimation is bypassed via PREDICT.XXX.

We now have a file named ALLCAST.USE which contains all the forecasts. The amount of time eliminated to develop the forecast versus just modeling and forecasting can be as much as 30:1 reduction, but in most cases it will be 20:1, a significant time saving feature especially when the modeling process is complex (i.e.. regression equations, daily time series, etc.).

Similarly, if you wish to have Autobox batch optimize the model coefficients and then forecast put a '2' at the end of the line:

OUTPUTATCH1 WEI3 2

If you wish to both TUNE and RECONFIGURE the model then put a '3' at the end of the line:

OUTPUTATCH WEI3 3

Chapter 4
Overriding the Expert System

Why would you need this? When you run Autobox using option "1" it uses specific default conditions within the algorithm. If you want to override the defaults, we let you do that.   You need to create trigger files in the installation directory and there is existence will cause Autobox to react. Some overrides require data to be in the file which are simple text files. The trigger files need to be named "*.AFS".

Here are the trigger files in alphabetical order:

2season.afs - THIS IS NOT JUST A TRIGGER FILE, BUT ALSO HAS CONTENT. If you have an unusual time series in that it doesn't have standard seasonality (ie monthly, quarterly, hourly) you can still model it using Autobox. If you have an hourly data that only covers 20 out of the 24 hours of the day, you can create a special file named '2season.afs" in the installation directory. Inside of the this text file place a '20' in it so that Autobox can model it.

AFS2SIP.afs – This is called "Simulated Forecasting" or bootstrapping via Monte Carlo methods. This comes with the Enterprise + version of Autobox. Autobox generates Sipmath 2.0 ready XML files(see probabilitymanagement.org for more on their free and very useful simulation tool). Put "1000" in this file and Autobox will generate 1,000 simulations for each of the forecasting periods(see image below). The second line can have a 0 or 1. This determines if outliers will be allowed to exist in the future or not. We recommend you put a 1 in the second line so that outliers that were identified and adjusted for in the history will be allowed to occur in the forecast and in effect widen the confidence limits providing more realistic version of uncertainty! If you put a 0 on the second line then then no outliers are used in the forecast(more for educational use to show you the old vs the new). We recommend using 1,000 so that there are enough samples/trials are drawn to provide good results. If you are forecasting out 36 months then you will have a total of 36,000 forecasts along with the average, standard deviation, min and max. After running, look for a file in the Autobox folder with an "XML" file extension (ie.Sales.XML). The sampling needs a "seed value" which we use the clock to drive from. If you run an example, you will not get the same results as the time of day changes. If you want to reproduce, you can create a file called "seed.afs" and put a number in it so that the results don't change. We report the seed at the bottom of details.htm that was used in the simulation which you can post to seed.afs if you want to use that same outcome. You need to delete nosim.afs from the Autobox folder as there is some additional compute time to do this step. 500 would work fine and 1,000 gives similar results. 1,000 is actually testing the limits(think possible crash?) of Autobox with all of

the memory that it uses with complicated algorithms.

Autobox generates *.FUM files(forecast univariate model) for each of the causal variables. Make sure to delete the previous *.FUM files if you made a change to the data.

<SLURP name="Example" coherent="True" about="Example SLURP of forecasts" origin="Autobox">												
<SIP name="Forecast00001" Base=" 58324. " SipMean=" 58296. " SipStd=" 596.99 " SipMin=" 57140. " SipMax=" 59326. " type="CSV" count=" 1000" csvr="02" about="Example SIP">												
58127.00, 58221.00, 58576.00, 58565.00, 57532.46, 58324.18, 59326.46, 58216.92, 57947.00, 58127.00, 58221.00, 57532.46, 58979.00, 57436.												
</SIP>												
<SIP name="Forecast00002" Base=" 55810. " SipMean=" 55783. " SipStd=" 599.30 " SipMin=" 54626. " SipMax=" 56812. " type="CSV" count=" 1000" csvr="02" about="Example SIP">												
55702.92, 55433.00, 55613.00, 55707.00, 55018.46, 56465.00, 54922.00, 55925.00, 56170.50, 55810.18, 54922.00, 56074.92, 55810.18, 55810.												
</SIP>												
<SIP name="Forecast00003" Base=" 58451. " SipMean=" 58437. " SipStd=" 602.39 " SipMin=" 57267. " SipMax=" 59453. " type="CSV" count=" 1000" csvr="02" about="Example SIP">												
58566.00, 58811.50, 58451.18, 57563.00, 58715.92, 58451.18, 58451.18, 57358.00, 57659.46, 59426.50, 58074.00, 59106.00, 59106.00, 58625.												
</SIP>												
<SIP name="Forecast00004" Base=" 59824. " SipMean=" 59807. " SipStd=" 603.81 " SipMin=" 58640. " SipMax=" 60826. " type="CSV"												

If retained data is in effect then simulation is disabled
If # of forecasts is greater than the maximum allowed # of series then simulation is disabled.
If intermittent demand is in effect then simulation is disabled (note nointerm.afs could control for this)
If non-constant variance is detected (tsay test) then simulation is disabled is disabled (note notsay.afs could control for this)
If power transform is in effect then simulation is disabled (note novarcon.afs could control for this)
If maximum # of allowed series is less than 150 (enterprise or Pro versions) then then causal simulation is disabled
Causal simulation is only available for type "0" series or type "2" series
Be careful with causalsin that if you run and Autobox

This suppresses the breakin.txt and breakout.txt reports which detail how the fit and forecast are generated. To suppress simulation create n override called nosim.afs.

Foreconf.afs – THIS IS NOT JUST A TRIGGER FILE, BUT ALSO HAS CONTENT. This determines the confidence level for the confidence limits (For 80% confidence limits put '80.0')

Integer.afs – Converts forecasts to integers

Interp.afs – Uses Interpolation to correct missing data.

Makealli.afs – Keeps all causal variables in the model.

Noarima.afs – Suppresses the use of ARIMA modeling. With daily data you don't want to overreact to the most recent data too much as the level shifts will do that for you.

Daily data relies on monthly and daily dummies (ie deterministic) behavior.

Nofixday – For daily data, If you don't have special days of the month.

Noholdum.afs - If a user specifies their own holidays this suppresses Autobox from automatically generating holiday variables (ie only matters when series name is "__010108Y11" with daily data)

Nointer.afs – Stops testing for interactions. For use with unusual seasonalities only.

Nointerm.afs – Stops looking for intermittent demand. Useful with daily data

Nolevel.afs – Stops testing for level shifts

Noparcon.afs – Stops the testing & adjustment for constancy of parameters (ie. Removes older data that has a different pattern than the more recent data)

Nopulse.afs - Stops testing for pulse outliers which is useful when you want to believe the outliers

Norule.afs – To stop these two overrides use this file. For daily data, Autobox will check if all weekends are zero and force the forecast weekends to be zero and for holidays the day of the week is not included to avoid double counting.

Noseasp.afs - Stops testing for seasonal pulse

Nostepdn.afs - Stops removal of causals.

Notrend.afs - Stops testing for local time trends.

Novarcon.afs – Stops the testing & adjustment for constancy of variance (ie. weight the observations based on their variance)

Noweekly.afs – When you want to force monthly dummies for short daily data. FYI: Autobox flips between monthly dummies and weekly dummies based on the length of the series.

Numbfore.afs – THIS IS NOT JUST A TRIGGER FILE, BUT ALSO HAS CONTENT. This defines the number of forecasts (For 24 forecasts put a '24' in this file)

Positive.afs – Converts forecasts to positive values

Sparse.afs – Specify two digits (ie .25 for 25%) that govern intermittent demand. Specify the % of zeroes in a dataset that trigger. Default is 25% in Autobox.

Parpharm.afs - Put a '1' for a mean model. Refer to the Appendix "Creating your ASC file" and you will see that "Objective 1" has a list of different prespecified models that can be used to override Autobox.

Stepupde.afs – THIS IS NOT JUST A TRIGGER FILE, BUT ALSO HAS CONTENT. This defines the number of interventions to be used in the model. (For 5 interventions put a '5' in this file) for daily data we recommend 80 outliers for 3 years of data.

Weekinmo.afs – Creates weekly dummy effects – Good for daily data that has a big month end, for example.

Here's a trick when you have long lagged promotion response or "Adstock" type of data. If you specify the causal series name with the words "DYN14" Autobox will react by modeling the promotion to decay over the next 14 periods is in effect. Note that the data type MUST be '3' for this to work. If you recall, a Data Type of 3 is: **"Future values are user specified; contemporaneous and lag and lead effects allowed"**. Also, if you have a "patch of outliers" that are in the same range (ie all zeroes) then you can use a '1' indicator during that patch. If you have a "patch of outliers" that varies wildly (ie high, low, etc.) then use "DYN" and the length and again data type must be equal to '3'.   You can put a name after the DYN to describe it so "DYN14_promo" would work.
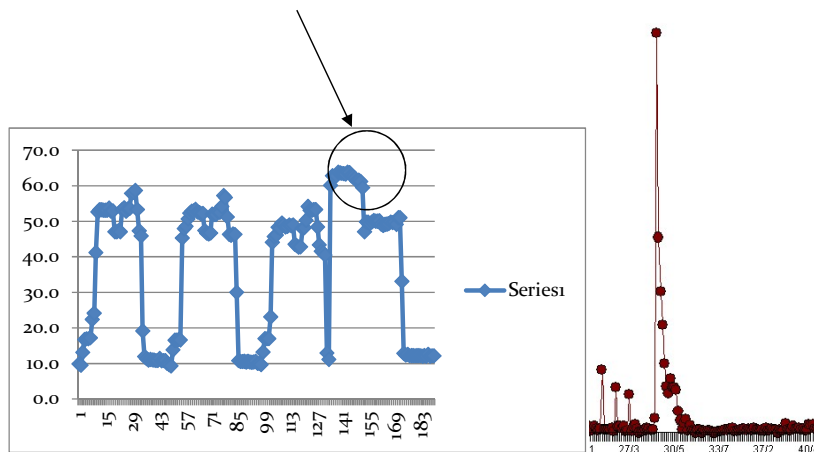
Fig. 5-0 – A Forecast with a Patch of Outliers Being Shown

Chapter 5
More In-Depth look at the Model and Rules Wizard

As the wizard indicates, your first need to choose what you want to do.

Option 1 – Run using proprietary AFS rules
Option 2 – Run using User's overrides of AFS rules
Option 3 – Choose a specific model like Holt Winters – it's only there for comparison in our opinion!
Option 4 – Choose your own model – this is for experts...be careful here
Option 5 – Runs only the identification process

Option 6 – Simulation - Create a series based on a model

If you select option '2' and press "Next", you will be given the option to use the "Edit/Use Present or Existing Rules" (press the down arrow to be shown what is available). This will use the current rules or you select rules you had saved previously. If you want to reset the conditions back to the original factory installed defaults used choose "Restore Default Rules" and click "Next".
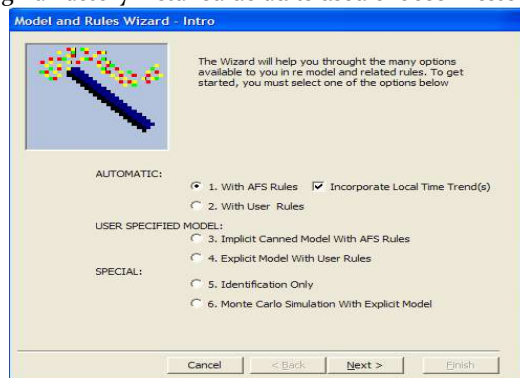
Fig. 5-1 – Rules Wizard
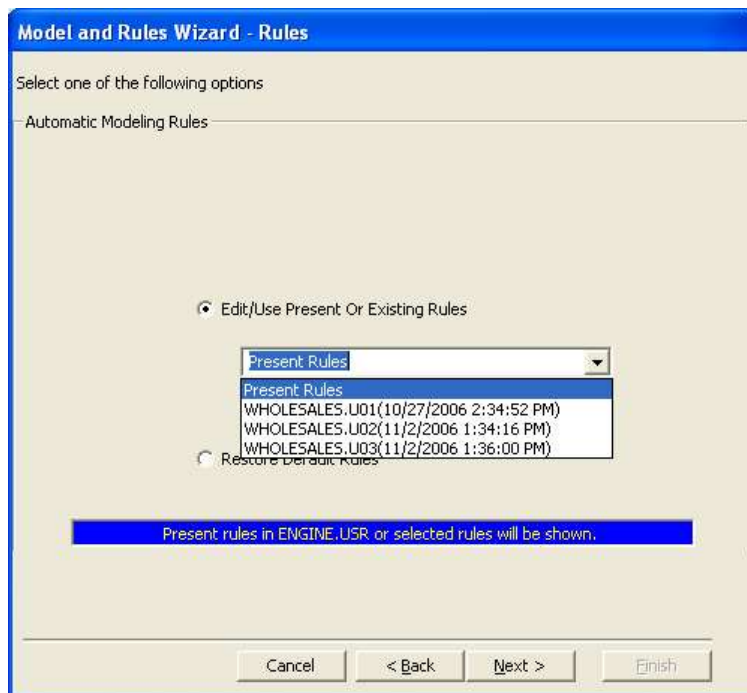
Fig. 5-2 – Option 2, Selecting Present Rules

When you select the rules and press "Next", they will be shown in the following order for your review and/or editing (please note that the rules vary between Noncausal(single) series and Causal(dependent series with independent series)series –we will show the causal screens after this section.

If you select this option, Autobox will check to see if more than 25% of the observations are zero and then run the following scheme to predict the intermittent demand.

Two new time series will be created: the interval and the rate. The interval series will be the number of periods with zero demand between periods where there is demand (i.e. If there is demand at every time period then the interval would be 1). The rate is the interval divided by the demand. We find this as a better alternative to Croston's method.

Autobox approaches the case of Intermittent or Sparse Demand by making the problem into a causal problem. Autobox tracks the count of how many periods were zero and creates an "interval" time series. It takes the volume and divides it by the interval to get a "rate". It projects out the interval and rate into the future and divides the two to get a forecast. Autobox will do all of this while still searching for level shifts. You need to have >25% zeroes in the series and 4 intervals to trigger intermittent demand.

You can override the % by creating a file in the installation directory named sparse.afs with .70 in the file to reflect a threshold of >70%.

You can also suppress intermittent demand by creating a file called "nointerm.afs" in the Autobox folder. We recommend creating an override file called "nopulse.afs" to NOT search for pulse outliers when dealing with intermittent data. We allow you decide which way to run it.

Fig. 5-3 – Option 2, Determining the Model

What are the Odds of Making the Month End Number?

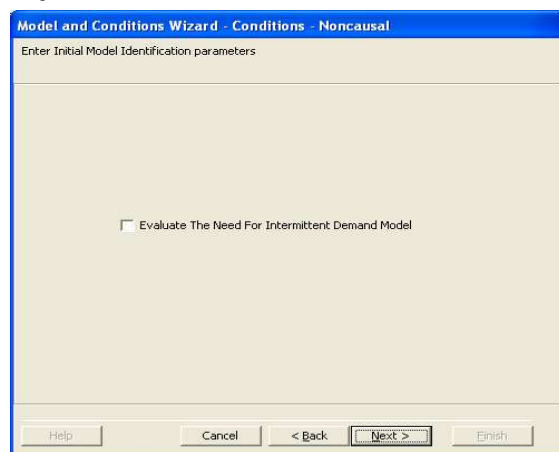Everyone wants to know as quickly as possible if the month end number is in jeopardy. If you have daily data and the number of forecasts is equal to the remaining days in the month, Autobox will report a file with a "CAN" extension that gives various probabilities to assess the likelihood of making various month end sums. You can easily interpolate to locate your goal and locate the probability. Make sure you use afs2sip.afs when using this analysis(search on afs2sip.afs)

Since days of the week vary, you can't use a simple ratio estimate to forecast the month end. You need to use sophisticated tools like Autobox to account for trend, day of the week variations, etc to get a robust forecast. The file canwe.asc comes with the installation will create canwe.csv:

PROBABILITY OF	EXCEEDING TARGET	__020910a
100	7617.881	__020910a
100	7657.477	__020910a
100	7697.073	__020910a
100	7736.669	__020910a
100	7776.265	__020910a
100	7815.862	__020910a
100	7855.458	__020910a
99.999	7895.054	__020910a
99.999	7934.65	__020910a
99.997	7974.247	__020910a
99.994	8013.843	__020910a
99.987	8053.439	__020910a
99.974	8093.035	__020910a
99.95	8132.631	__020910a
99.905	8172.228	__020910a
99.825	8211.824	__020910a
99.687	8251.42	__020910a
99.455	8291.016	__020910a
99.077	8330.613	__020910a
98.482	8370.209	__020910a
97.574	8409.805	__020910a
96.239	8449.401	__020910a
94.345	8488.997	__020910a
91.752	8528.594	__020910a
88.336	8568.19	__020910a
84	8607.786	__020910a
78.709	8647.382	__020910a
72.499	8686.979	__020910a
65.495	8726.575	__020910a
57.903	8766.171	__020910a
57.903	8766.171	__020910a
50	8805.767	__020910a
42.097	8845.364	__020910a
34.505	8884.96	__020910a
27.501	8924.556	__020910a
21.291	8964.152	__020910a
16	9003.748	__020910a
11.664	9043.345	__020910a
8.248	9082.941	__020910a
5.655	9122.537	__020910a
3.761	9162.133	__020910a
2.426	9201.73	020910a

Fig. 7-2 – canwe.csv

The 50th percentile is all of the forecasts for the rest of the month summed up from the Autobox generated forecast.

You can locate your month end number probability on the table by using interpolation.

Let's say that the month end number is 7,770 then you can say that there is a 100% chance of making the number (barring any major intervention!)

Safety Stock

Autobox provides a safety stock calculation at a 90% service level out for as many periods as you are forecasting. Note that if there are not enough historical observations and given the seasonality and length of the forecasts this report won't be produced. The file created is called "SAFETY.TXT".

For an 8 week forecast, we compute the mean square error of the fit and then we generate forecast variances for the next 8 weeks. We then compute the covariance of the forecast variances and use them to create a forecast of the sum for each of the 8 weeks.

```
        SERVICE LEVEL    :  90 %

CYCLE          VARIANCE OF THE SUM   STD DEV OF THE SUM      SAFETY STOCK
  1                    0                    1                     1
  2                    1                    1                     1
  3                    4                    2                     3
  4                    5                    2                     3
  5                    5                    2                     3
  6                    9                    3                     4
  7                   10                    3                     4
  8                   13                    4                     5
  9                   13                    4                     5
 10                   19                    4                     5
 11                   20                    4                     5
 12                   20                    4                     5
 13                   23                    5                     6
 14                   23                    5                     6
 15                   24                    5                     6
 16                   26                    5                     6
 17                   28                    5                     6
 18                   29                    5                     6
 19                   29                    5                     6
 20                   34                    6                     8
 21                   34                    6                     8
 22                   36                    6                     8
 23                   36                    6                     8
 24                   37                    6                     8
 25                   40                    6                     8
 26                   40                    6                     8
 27                   44                    7                     9
 28                   45                    7                     9
 29                   48                    7                     9
 30                   48                    7                     9
 31                   55                    7                     9
 32                   55                    7                     9
 33                   56                    7                     9
 34                   59                    8                    10
 35                   59                    8                    10
 36                   60                    8                    10
```

Fig. 7-3 – Safety Stock Table

We will skip "Multiple Groups/Panels for now. It is discussed later. If you choose to change the "convergence criteria", you really won't need to make a change to the defaults. You can change the way Autobox iteratively goes about identifying the model. The % change in error sum of squares is set to 0 so you can't be more precise. The % change in parameter values are set to .1% and you can increase the iterations to a maximum of 200, but it doesn't make much of a difference.
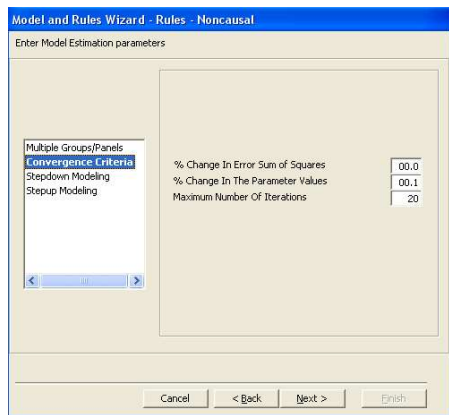

Fig. 5-4 – Option 2, Entering Model Parameters

For step down modeling (where variables are removed or weighted), you can either adjust the confidence level for the option or "uncheck" the box so that the test is disabled. If these are checked then:

Necessity - This will drop all variables that are introduced by Autobox (i.e. outliers). (The opposite case is that it keeps all variables)

Test for Constancy of Parameters - Autobox will determine if you have too much historical data that does not match the more recent data and delete older data.
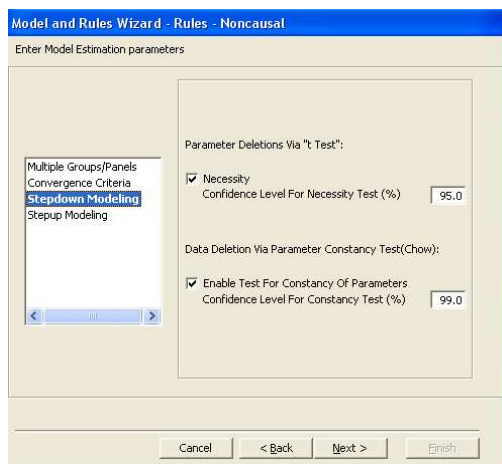
Fig. 5-5 – Stepdown Modeling Parameters

For stepup modeling (where variables are added into the model by Autobox), you can either adjust the confidence level for the option, "check" or "uncheck" the box, or modify the counts.

Sufficiency Test for Stochastic Structure- This will add in ARIMA variables if significant.

Sufficiency Test for Determinstic Structure – This will add in Intervention variables.

Maximum No. Of Outliers To Be Identified – This will limit the number of outliers that can be introduced.

Include Pulse Variables – This will allow one-time unusual variables to be accounted for.

Include Step Variables – This will allow sudden changes in mean to be accounted for.

Min. no. Of Observations In Group – This determines how many observations you need before you can create a step variable.

Include Local Trends – This looks for changes in intercept.

Include Seasonal Pulse Variables – This looks for pulses that occur every season.

Enable Auto Fix Up For Fixed Effects – For example, for monthly data, Autobox would include 11 dummies for monthly data to account for fixed effects.

Discrete Change Test For Variance – This will search for changes in variance and weight the history based on the observations.

Minimum number of Residuals to Pool – There are two groups created to compare the variance and this determines the smallest size of one of the groups.

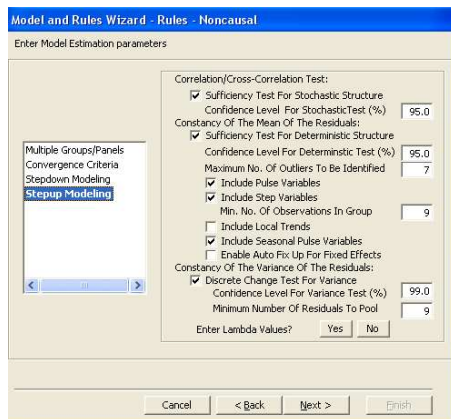**Enter Lamda Values –** click "Yes" to search for optimal variance stabilization

Fig. 5-6 – StepUp Modeling Parameters

Several common used variance-stabilizing transformations

Relationship of σ^2 to $E(y)$	Transformation
$\sigma^2 \propto$ constant	$y' = y$ (no transformation)
$\sigma^2 \propto E(y)$	$y' = \sqrt{y}$ (square root: Poisson data)
$\sigma^2 \propto E(y)(1 - E(y))$	$y' = \sin^{-1}(\sqrt{y})$ (arcsin)
$\sigma^2 \propto (E(y))^2$	$y' = \log y$
$\sigma^2 \propto (E(y))^3$	$y' = y^{-1/2}$ (reciprocal square root)
$\sigma^2 \propto (E(y))^4$	$y' = y^{-1}$ (reciprocal)

The bottom option is the "Enter Lambda Values". This allows you to determine the form of the Box-Cox Transformation. The first observation is always 1 as a rule. Enter 0 for logs in row 2 and -1 for reciprocals in row 3, for example

Fig. 5-7 - Setting Lambda Values

Click "Next" and the "Model Forecasting" parameters will be shown as follows(Please note that to get an explanation/definition of a particular checkbox or textbox merely click on it and press F1):

Fig. 5-8 – Model Forecasting Parameters

This is the process of using the model that best describes the past observations (to extrapolate the pattern of the data) to predict Confidence Levels into the future.

If this option is not enabled, then no forecasts will be made.

Press "Next" and you will be given the option to save changed rule and/or the final model as indicated in the following:
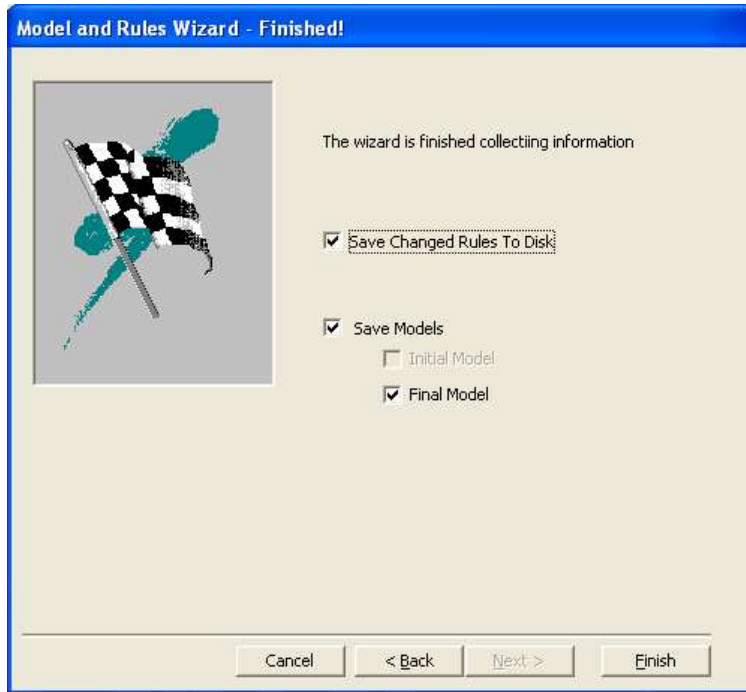

Fig. 5-9 – Model and Rules are Set, Prepared to Save

Again, press "Finish" and the dataset will be processed automatically.

If you selected option '2' and had causal variables then this additional screen would have been presented to you at the beginning.

Use Difference Factors From ARIMA Model In Causal Model – This might be helpful in identification and problematic in estimation.

Constrain All User Causal Coefficients in Model – This keeps all user causal variables in model regardless of their significance.
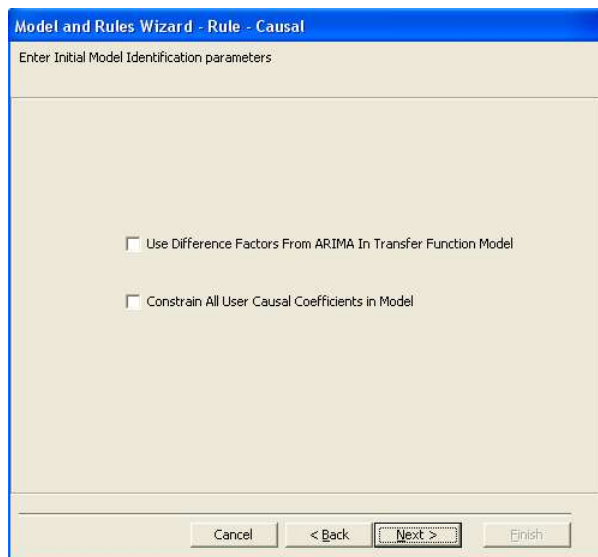

Fig. 5-10 – Causal Rules

If you select option '3' and press "Next", you must select one of the offered models as indicated in the following:
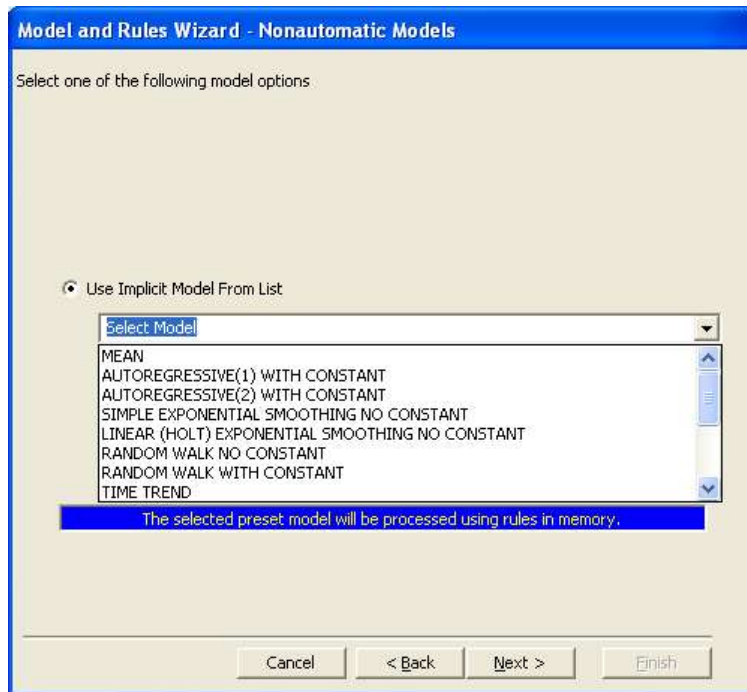

Fig. 5-11 – Option 3, Models

If you click the box "Allow Modifications" then Autobox will use all of its "expert system" approaches to try and create a better model given you picked a starting model like "Time Trend" as seen here. Note that we believe this is dangerous territory as all of the models in this area are typically a bad place to start.
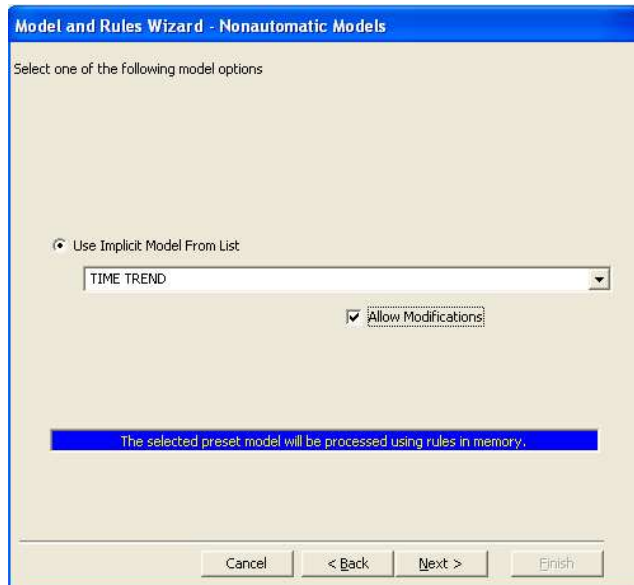

Fig. 5-12 – Option 3, Selecting a Preset Model

Press "Next" and you will have the option to save the final rules and model, then press "Finish" to process the dataset.

If you have causals then you have two "pre-selected" models to choose from along with the ability to choose "Allow Modifications". Again, this is dangerous territory as opposed to using option '1'.
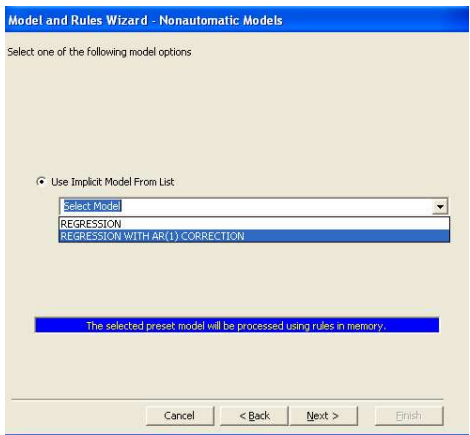

Fig. 5-13 – Option 3, Modifications

If you select option '4' and press "Next", you can choose to make a "New" model or use one of the saved models indicated in the following. Click "Next"
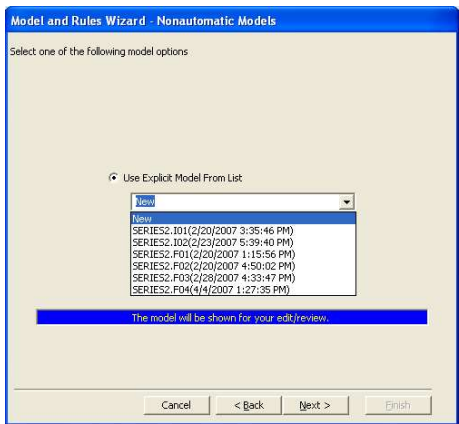

Fig. 5-14 – Option 4, Set Model

If you only have one series in your dataset you see the words "Noncausal" twice at the top of the screen.

If you only have multiple series(not shown) in your dataset you see the words "Causal" at the top and the word "noise" here. The noise refers to the ARIMA structure in the transfer function model. We will show you a causal example on the next screen.

Fig. 5-15 – Option 4, Noncausal Model Creation

Here we specify the model for aom73, but we click "On Copy Parameters To" and then click on all of the other causal series so that they have the same model applied.

Fig. 5-16 – Option 4, Copy Parameters on All Series

Press "Next" and you will be given the option to use the "Edit/Use Present Or Existing Rules" (press the down arrow to be shown what is available) or "Restore Default Rules" and as follows:
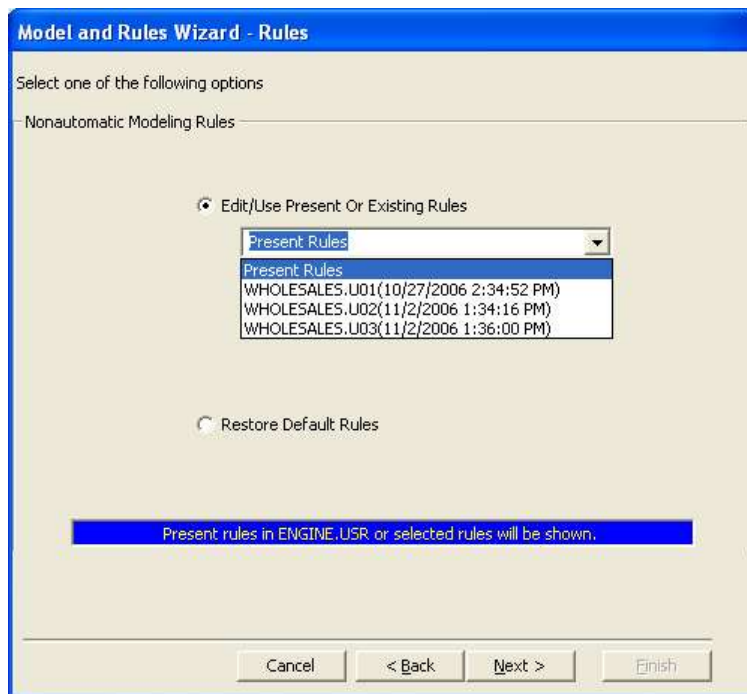
Fig. 5-17 – Option 4, Setting Rules

Press "Next" and the rules will be shown as follows for your review and/or edit:
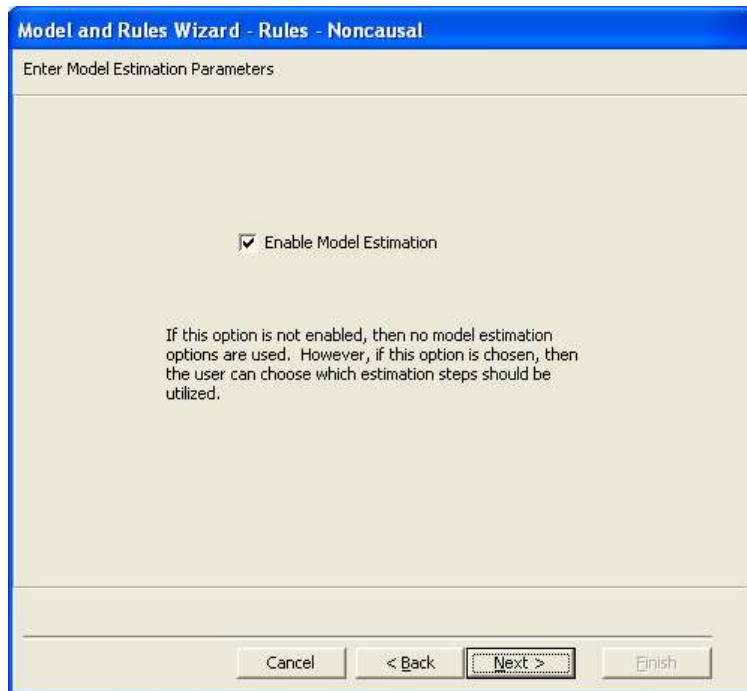
Fig. 5-18 – Option 4, Enabling Model Estimation

Estimation and diagnostic checking represent the second phase of the Box-Jenkins modeling procedure. The estimation option computes the model coefficients via nonlinear least squares and generates residuals. Various statistics are computed for both the estimated parameters and the residuals from the model.

Press "Next" and the "Model Forecasting" parameters will be shown as follows:

Fig. 5-19 – Option 4, Model Forecasting Parameters

This is the process of using the model that best describes the past observations (to extrapolate the pattern of the data) to predict Confidence Levels into the future.

If this option is not enabled, then no forecasts will be made.

Press "Next" and you may elect to save changed conditions and/or the "Initial Model" and/or the

"Final Model" as shown in the following"
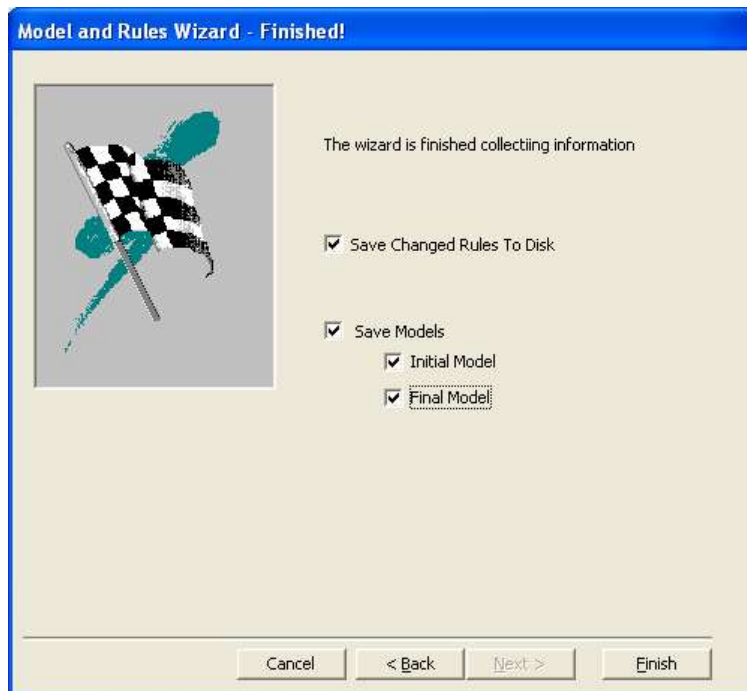
Fig. 5-20 – Option 4, Saving the Model

Again, press "Finish" to process the dataset.

If you select option5, press "Next" then "Finish" to process the dataset. Autobox will only perform identification. This is good for beginners just trying to learn Box-Jenkins to see if they can identify the model on their own.

Press "Next" and you will be given the option to provide a "New" model or select a saved model as indicated in the following. You can create your own model:
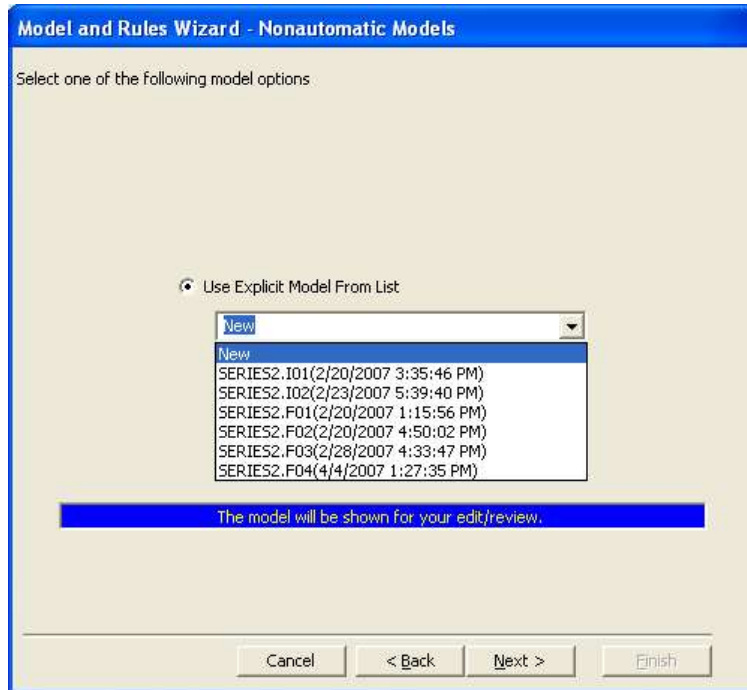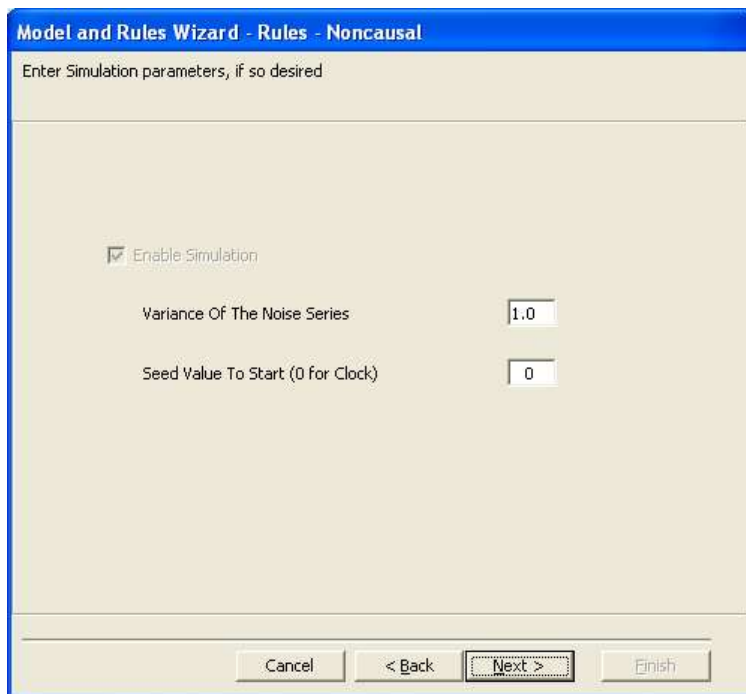
Fig. 5-21 – Option 5, Creating a Model

Press "Next" and you will be shown the model parameters for you review or edit. Press "Next" again and you will be given the option to edit the simulation parameters as follows:

Fig. 5-22 – Option 5, Editing Simulation Parameters

Press "Next" and you may elect to save the changed rules then press "Finish" to process the dataset.

Troubleshooting

If you get an error or are very dissatisfied with the forecast, model, etc.   We would like to receive an email with a zip file with the following:

Data file ("*.ASC")
*.AFS files
Autobox.exe, freefore.dll, freel.dll
 ENGINE.*,
Create a file named 'snoop.afs' and 'apush.afs' and run Autobox and save all files that were created in the Autobox directory during the process of running the ASC file (You can use Windows Explorer and sort the directory to see these newly created files). Open a Word document and take print screens of each of your selections up so we can see exactly what you did. Upon receipt of your email, we will review and respond. Make sure you delete snoop.afs and apush.afs as they are diagnostic files!

IF AUTOBOX STOPS WORKING (YOU CAN TELL THIS BY VIEWING DETAILS.HTM ON YOUR FREEFORE DIRECTORY THEN AUTOBOX IS NOT FUNCTIONING PROPERLY) AND YOU NEED TO STOP IT, DON'T TURN OFF YOUR COMPUTER AND DO FOLLOW THESE DIRECTIONS

> HIT CTRL-ALT-DEL AT THE SAME TIME
> CHOOSE "TASK MANAGER"
> CHOOSE "PROCESSES"
> SORT ON IMAGE NAME BY CLICKING ON IMAGE NAME
> SELECT AFSENGINE.EXE AND CLICK ON "END PROCESS"
> SELECT AFSLITE.EXE AND CLICK ON "END PROCESS"
> CHOOSE "APPLICATIONS"
> SELECT "AUTOBOX" AND CHOOSE "END TASK"

Contact us for any questions:

Afs Inc.

P.o. Box 563

Hatboro PA 19040

sales@autobox.com

Phone 215-675-0652   Fax 215-672-2534

APPENDIX

Creating your own .ASC file

The ASC file contains some header records and the time series. This is really an easy procedure; but the information and data must be entered in a text file in a very specific order in a single column. We would prefer to first show you an example of an ASC file. There is some text below this example that explains the options for each line in great detail. Call us if you are the slightest bit confused.

We highly recommend that you download/install our interactive version (http://www.Autobox.com/abox.exe) as a way of Quality Control on how you have built the ASC file. You can quickly establish if there are any problems by opening the ASC file that you have built.

0
0
0

> These are the three "Objectives". We prefer to leave these options all at zero, but you can change these as you wish. They determine automatic modeling, how much output to show

1
12
1949
1
144
0
0
12
BJ07

> These are the 8 "Data Properties". They define the characteristics of the series like the seasonality, beginning year, beginning period, number of observations.

> This is where the name of the series is specified. This field is "variable" in that if you have a causal problem where you have many series then you would list all of the names of the series HERE BEFORE the data. YOU should not USE THE SAME SERIES NAME FOR A DEPENDENT SERIES IN TWO SEPARATE ASC FILES as the output from a modeling run will overwrite each other. Also, by specifying the date (January 1, 2006) with the name in this format "__010106Y11" and line 5 has a '7' then day of the week effect is analyzed

0
112.000000
118.000000
132.000000
129.000000
121.000000
135.000000
148.000000
148.000000
136.000000
119.000000
104.000000
118.000000

> This is the famous Box-Jenkins airline series (abridged) with the data shown here. If you have a causal problem then each of the series are placed end-to-end downward in a Blocked rectangular historical array. The last series would be the dependent series.
>
> BUT If you have future values, you place this future values series below the dependent series
>
> BUT If you have retained values, you place this series below the future values series
>
>   Output series
>
>   Future
>
>   Retained values

This is a summ

Objectives (all are required)

Data properties (all are required)

Data names (in the order of 1st input series to nth input series, if any; and then the output series)

Data type (in the same order as the data names)

Historical Data (in the same order as the data names)

Future Values (for all input series which have a data type of 1, 2, or 3, if any, in the same order as data names)

Retained Data (If any, in the same order as the data names)

The following structure tables indicate the parameters and/or limitations for each of the above categories. OBJECTIVE Structure:

Name	Description
OBJECTIVE(1)	Sets forth the model conditions as indicated by the following: 0 = Totally Automatic NONCAUSAL MODELS IN AUTOBOX BATCH MEMORY 1 = MEAN 2 = AUTOREGRESSIVE (1) WITH CONSTANT 3 = AUTOREGRESSIVE(2) WITH CONSTANT 4 = SIMPLE EXPONENTIAL SMOOTHING NO CONSTANT 5 = LINEAR (HOLT) EXPONENTIAL SMOOTHING NO CONSTANT 6 = RANDOM WALK NO CONSTANT 7 = RANDOM WALK WITH CONSTANT 8 = TIME TREND 9 = TIME TREND PLUS AR(1) CORRECTION 10 = FOURIER 11 = HOLT LINEAR TREND PLUS ADDITIVE SEASONAL FACTORS (TREND FORM) 12 = DAMPED TREND LINEAR EXPONENTIAL SMOOTHING NO CONSTANT 13= SEASONAL EXPONENTIAL SMOOTHING NO CONSTANT 14 = HOLT LINEAR TREND PLUS ADDITIVE SEASONAL FACTORS (ARIMA FORM) 91 = FIND HIDDDEN SEASONALITY THEN SET TO "BEST" 97 = IDENTIFICATION ONLY 98 = HOLT-WINTERS TREND PLUS MULTIPLICATIVE SEASONAL FACTORS (TREND FORM) CAUSAL MODELS IN AUTOBOX BATCH MEMORY 51 = REGRESSION 52 = REGRESSION WITH AR(1) CORRECTION 53 = STEPDOWN REGRESSION 54 = STEPDOWN REGRESSION WITH AR(1) CORRECTION STARTING MODEL SUPPLIED: 99 = STARTMOD.123 199 = STARTMOD.123 + SIM IF STARTING MODEL SUPPLIED = 99 or 199, STARTMOD.123 is required. 200 = TOTALLY AUTOMATIC + ABOXLITE model is developed

OBJECTIVE (2)	No longer used leave as a 'o'
OBJECTIVE (3)	No longer used leave as a 'o'

DATAPROP Structure:

Name	Description
DATAPROP(1)	Number of series in the problem
DATAPROP(2)	Seasonality. How often the data was sampled. (i.e. 52 for weekly data) Please note that all series in the model must have the same seasonality
DATAPROP(3)	Beginning year. The year or major number identifying the starting point of the data. Please note that all series in the model must have the same Beginning Year. If you wish to use series whose original Beginning Year are different, you must determine the common matrix for the series and use that starting point as the Beginning Year.
DATAPROP(4)	Beginning period. the starting point of the data. (i.e. 1 for the 1st week in the year) Please note that all series in the model must have the same Beginning Period. If you wish to use series whose original Beginning Period are different, you must determine the common matrix for the series and use that starting point as the Beginning Period.
DATAPROP(5)	Number of historical values in each of the time series in the model
DATAPROP(6)	Number of future values to be included for each applicable input series. If a causal model (includes a dependent and independent series) and the DATATYPE of any the input(independent) series is 1, 2, or 3, enter DATAPROP(7) + Number of Future Values (this must equal the number of forecasts to be calculated) to be supplied by the user. If DATATYPE of all input series is 0, or if a noncausal model, this must show a 0.
DATAPROP(7)	The number of values retained from the end of the series to be used to evaluate prior forecasts (enter 0 if none)
DATAPROP(8)	Number of forecast values to be calculated

You should have a row for each series name

DATANAME Structure:

Name	Description
DATANAME	Actual name of each series in model in the order 1st Input series, 2d input series, …N input series, output series
	These names must be limited to 22 characters for Input series and 14 characters for the output series; and they cannot contain space(s), period(.), exclamation point(!), backquote('), brackets([]), wild card characters such as * or ?, and control characters(ASCII values 0 through 31).
	Also, by specifying the date (January 1, 2006) with the name in this format "__010106Y11" and line 5 has a '5' or a '7' then daily effects such as "day of the week", "week of the year", "day of the month" are analyzed. Note:Day of the month only applies for '7'.

You should have a row for each series

DATATYPE Structure:

Name	Description
DATATYPE	Integer value for series type. This can be 1 of 4 values
	0 = Future Values are self-projected; contemporaneous and lag effects allowed. All output series must be 0.
	1 = Future Values are user specified; contemporaneous effect allowed.
	2 = Future values are user specified; contemporaneous and lag effects allowed .
	3 = Future values are user specified; contemporaneous, lag and lead effects allowed.
	A '0' tells Autobox you want it to forecast future values of the causal. Types '1','2', and '3' are user supplied future values. In terms of effects on run-time Type '3' would take the longest, Type '1' takes the shortest and '2' and '3' would be similar.

The following is an example of an .ASC file for a noncausal (single) series{annotations are not included in the file}:

0	(objective(1) indicates totally automatic modeling)
0	(objective(2) indicates use default conditions in memory)
0	(objective(3) indicates full output)
1	(DataProp(1) number of series in the problem set)
52	(DataProp(2) seasonality of the series)
1998	(DataProp(3) beginning year or major period)
2	(DataProp(4) beginnng or minor period)
67	(Dataprop(5) number of historical data in series)
0	(DataProp(6) number of future values)
0	(DataProp(7) number of retained data)
24	(DataProp(8) number of forecasts to be calculated)
pizza	(output series name)
0	(data type)
15	(historical data – 67 observations)
14	
6	
.	
.	
12	

How to Specify Your Own Model

This is where an expert can use his expertise.

If you select option '4' and press "Next", you can choose to make a "New" model or use one of the saved models indicated in the following:

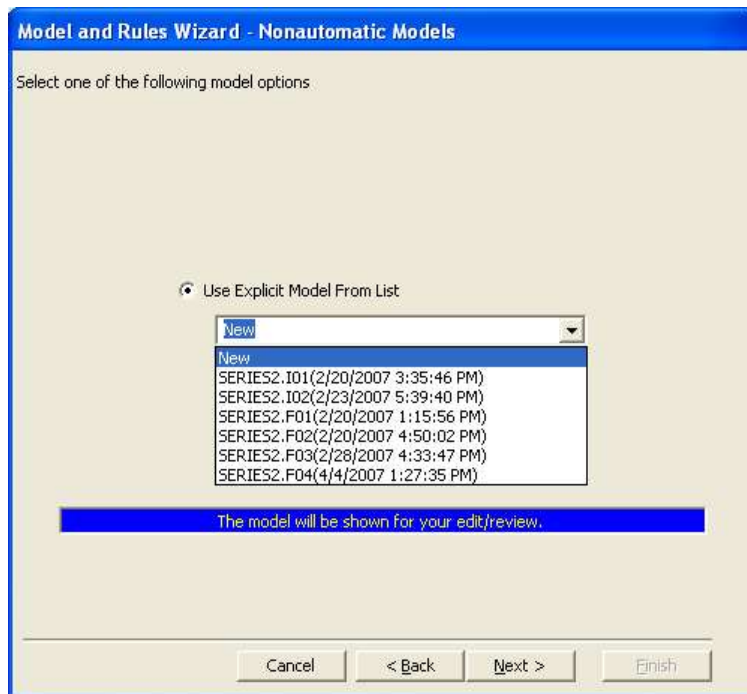**Fig. 3-6 – Selecting Model from Dropdown List**

Press "Next" and the model parameters will be shown for your review and/or edit as

follows:

Fig. 3-7 – Setting the Parameters for the Model

Please note that for Causal series we will show model parameters for each of the input (independent) series as well as the output(dependent) series. You can copy that same model to the other input variables.

Daily Data – Using Holidays outside of the U.S>

If you want to specify a holiday on your own, make sure to start the series name with a "M_" as Autobox has a useful check of the history for all years having a zero to override the forecast to be a 0.

Autobox incorporates U.S. Holidays into the modeling and forecasting process. If you have data from outside of the U.S., create a file called "noholdum.afs" to suppress the search for U.S. in the Autobox folder.

The TOP GDP producing countries were selected and there are 16 Countries that we have set up which are ready to be used in Autobox from 2001 out to the year 2027. If your forecast is only out to 2020, there is no need to adjust. To apply India holidays, create a file called "country.afs" and put the code as seen below (ie 5 for India) into a file called "country.afs" and save it to the Autobox folder. Autobox will use the Holiday variables from that Country. We will explain how you can customize Autobox for a different country not listed or even perhaps a local area within a country which follows a more specific holiday list than what we have included.

Code	Country
1	USA
2	BRAZIL
3	FRANCE
4	GERMANY
5	INDIA
6	ISRAEL
7	ITALY
8	JAPAN
9	MEXICO
10	RUSSIA
11	RWA
12	SOUTH AFRICA
13	SOUTH KOREA
14	SPAIN
15	SWITZERLAND
16	AUSTRALIA
17	UNITED KINGDOM

If you want to créate your own holiday file for a specific country you can use another country's holiday file as a template. For example, open the file "ROS01.HLD" found in your Autobox folder (which is the USA) and make changes to it and save it as ROS18.HLD. You can then créate a file called "country.afs" with an 18 and Autobox will use that to bring in the custom holidays.

Let's discuss the file format. We will not show you the whole file as it is too large.

The 18 in the first row tells Autobox how many holidays exist. If your country has only 12 then change this number to 12 and then save this file with a new name with a number higher than the ones listed on the page before(ie ros17.hld).   Row two will have the first holiday's name and a count of 27. The 27 represents that there are 27 years of this holiday in the file.   The date in the next row, 37250, would be the 3725oth day since 1/1/2001(our internal date reference point). The numbers below 27250 are increments of 365 and sometimes 366. Set row 3's date to match the holiday that you want to add and set the dates below it to match. Note: You don't need to set this up like we did to go out 27 years like we did. Just remember that you would need to change 27 down to a different number.

```
AXFBJR(  1)( 1:  ) ='18                    '
AXFBJR(  2)( 1:  ) ='M_CHRISTMAS          27'
AXFBJR(  3)( 1:  ) ='37250                 '
AXFBJR(  4)( 1:  ) ='37615                 '
AXFBJR(  5)( 1:  ) ='37980                 '
AXFBJR(  6)( 1:  ) ='38346                 '
AXFBJR(  7)( 1:  ) ='38711                 '
AXFBJR(  8)( 1:  ) ='39076                 '
AXFBJR(  9)( 1:  ) ='39441                 '
AXFBJR( 10)( 1:  ) ='39807                 '
AXFBJR( 11)( 1:  ) ='40172                 '
AXFBJR( 12)( 1:  ) ='40537                 '
AXFBJR( 13)( 1:  ) ='40902                 '
AXFBJR( 14)( 1:  ) ='41268                 '
AXFBJR( 15)( 1:  ) ='41633                 '
AXFBJR( 16)( 1:  ) ='41998                 '
```

```
AXFBJR( 17)( 1:  ) ='42363                    '
AXFBJR( 18)( 1:  ) ='42729                    '
AXFBJR( 19)( 1:  ) ='43094                    '
AXFBJR( 20)( 1:  ) ='43459                    '
AXFBJR( 21)( 1:  ) ='43824                    '
AXFBJR( 22)( 1:  ) ='44190                    '
AXFBJR( 23)( 1:  ) ='44555                    '
AXFBJR( 24)( 1:  ) ='44920                    '
AXFBJR( 25)( 1:  ) ='45285                    '
AXFBJR( 26)( 1:  ) ='45651                    '
AXFBJR( 27)( 1:  ) ='46016                    '
AXFBJR( 28)( 1:  ) ='46381                    '
AXFBJR( 29)( 1:  ) ='46746                    '
AXFBJR( 30)( 1:  ) ='M_CINCODEMAYO          27'
AXFBJR( 31)( 1:  ) ='37016                    '
```

Many rows removed to avoid redundancy

```
                    '

AXFBJR(503)( 1:  ) ='45972                    '
AXFBJR(504)( 1:  ) ='46337                    '
AXFBJR(505)( 1:  ) ='46702                    '
```

Change or make sure the numbers on the right are the counters for the first instance of a holiday. Note that Cinco De Mayo up above has a counter of 30 and 30 is in the second row shown below. Once you have done this, all you need to do is save the file and change the country.afs file and save it.

Chapter 6:
Engine.AFS, Engine. XXX, and FFLITE

ENGINE.AFS

Listing of ENGINE.XXX which is copied to ENGINE.AFS to micro-manage the dll

```
NAME OF FILE CONTAINING INITIAL MODEL          |STARTMOD.123
 MONTE CARLO SIMULATION ENABLED (0=NO 1=YES)    |  0
  VARIANCE OF THE NOISE SERIES           |01.0
  SEED VALUE TO START ( 0 FOR CLOCK )          |  0
NON-CAUSAL:                          |
 EVALUATE THE NEED FOR INTERMITTENT DEMAND MODEL |  0
CAUSAL:                              |
 MAXIMUM LEAD FOR SERIES WITH POSSIBLE LEAD EFFECT|  0
 USE DIFFERENCE FACTORS FROM ARIMA IN TF MODEL   |  0
 CONSTRAIN ALL USER CAUSAL COEFFICIENTS IN MODEL |  0
ENABLE MODEL ESTIMATION              |  1
 # OF GROUPS IN POOLED-CROSS SECT. T/S (IF ANY)  |  0
  SAMPLE SIZE IN EACH OF THE GROUPS         |
 % CHANGE IN ERROR SUM OF SQUARES         |00.0
 % CHANGE IN THE PARAMETER VALUES          |00.1
 MAXIMUM # OF ITERATIONS              | 20
 NECESSITY TEST:                 |  1
  CONFIDENCE LEVEL FOR NECESSITY           |95.0
 SUFFICIENCY TEST:(STOCHASTIC STRUCTURE )     |  1
  CONFIDENCE LEVEL FOR SUFFICIENCY (SS)       |95.0
 SUFFICIENCY TEST:(DETERMINISTIC STRUCTURE )   |  1
  CONFIDENCE LEVEL FOR SUFFICIENCY (DS)       |95.0
  MAXIMUM NUMBER OF OUTLIERS TO BE IDENTIFIED   |  7
  INCLUDE PULSE VARIABLES              |  1
  INCLUDE STEP VARIABLES              |  1
   MINIMUM NUMBER OF OBSERVATIONS IN GROUP     |  9
  INCLUDE SEASONAL PULSE VARIABLES         |  1
  INCLUDE LOCAL TRENDS              |  0
  ENABLE AUTOMATIC FIXUP FOR FIXED EFFECTS     |  0
  NUMBER OF LAMBDA VALUES TO EVALUATE IN EST    |  1
   LAMBDA VALUES TO EVALUATE (4F4.0)         |1.00
  DISCRETE CHANGE TEST FOR VARIANCE         |  1
   CONFIDENCE LEVEL FOR VARIANCE TEST        |99.0
```

```
  MINIMUM NUMBER OF OBSERVATIONS IN GROUP      |  5
CONSTANCY OF PARAMETERS:                   |   1
 CONFIDENCE LEVEL FOR CONSTANCY TEST          |99.0
STORE MODEL FORM (YES/NO)               |   0
 DISPLAY MANAGEMENT ANALYSIS              |   0
ENABLE MODEL FORECASTING               |   1
 # OF FORECAST VALUES TO COMPUTE          |  13
 CONFIDENCE LIMIT FOR THE FORECASTS (%)        |95.0
 CONVERT THE FORECAST VALUES TO POSITIVE VALUES  |  0
 CONVERT THE FORECAST VALUES TO INTEGERS       |  0
 CONVERT PULSE AT LAST OBSERVATION TO STEP      |  0
 CONVERT PULSE TO SEASONAL PULSE          |  0
DISPLAY IDENTIFICATION INFORMATION          |  1
 DISPLAY ACF TABLE DURING IDENTIFICATION       |  1
 DISPLAY ACF GRAPH DURING IDENTIFICATION       |  1
 DISPLAY CCF TABLE DURING IDENTIFICATION       |  1
 DISPLAY CCF GRAPH DURING IDENTIFICATION       |  1
 DISPLAY PREWHITENING MODEL(S)            |  1
 DISPLAY PREWHITENING MODEL(S) EQUATION        |  1
 DISPLAY IDENTIFIED MODEL              |  1
 DISPLAY IDENTIFIED MODEL EQUATION          |  1
DISPLAY ESTIMATION INFORMATION            |  1
 DISPLAY ESTIMATED MODEL PARAMETERS         |  1
 DISPLAY ESTIMATED MODEL EQUATION          |  1
 DISPLAY FIT VS ACTUAL TABLE             |  1
DISPLAY DIAGNOSTIC CHECKING INFORMATION       |  1
 DISPLAY RESIDUAL ACF TABLE             |  1
 DISPLAY RESIDUAL ACF GRAPH             |  1
 DISPLAY RESIDUAL CCF TABLE             |  1
 DISPLAY RESIDUAL CCF GRAPH             |  1
 DISPLAY NECESSITY TEST RESULTS           |  1
 DISPLAY SUFFICIENCY TEST RESULTS          |  1
 DISPLAY VARIANCE STABILITY TEST RESULTS       |  1
 DISPLAY OUTLIER TEST RESULTS            |  1
 DISPLAY CONSTANCY TEST               |  1
DISPLAY FORECASTING INFORMATION           |  1
 DISPLAY FORECAST MODEL PARAMETERS         |  1
 DISPLAY FORECAST MODEL EQUATION          |  1
 DISPLAY TABLE OF TRANSFORMED FORECAST VALUES   |  1
 DISPLAY TABLE OF FORECAST VALUES          |  1
```

INTRODUCTION TO THE ENGINE.XXX FILE

HOW THE ENGINE VERSION WORKS

The Autobox batch version uses internal choices to model. If you would like to customize some of the choices, you can do so here by Editing a file named "Engine.xxx" and saving it as "Engine.go". After the batch process runs it renames it as "Engine.afs" and will continue to use those overrides as long as "Engine.afs" exists. You can rename it to "Engine.old" if you want to stop using your overrides. The information seen to the right of the "|" character is where Autobox batch searches during an execution. A '0' means not used and a '1' means it is used.

Line 1 – Do not change this line

NAME OF FILE CONTAINING INITIAL MODEL |STARTMOD.123

This points Autobox batch to the file use as the starting model. This is all handled in the "BAT" file process for you.

Line 2

MONTE CARLO SIMULATION ENABLED (0=NO 1=YES) |0

Choose "1" to have Autobox batch create a time series based on a user specified model.

Line 3 - OBSOLETE

VARIANCE OF THE NOISE SERIES |1.0

Line 4 - OBSOLETE

SEED VALUE TO START (0 FOR CLOCK) |0

Line 5 - Title only

NON-CAUSAL:
Line 6

EVALUATION THE NEED FOR INTERMITTENT DEMAND |0

If a 0 exists then the series will not be considered for intermittent demand modeling. If a 1 exists then if >25% of the data has a zero then intermittent demand modeling will proceed.

Line 7 - Title only

CAUSAL:

Line 8

MAXIMUM LEAD FOR SERIES WITH POSSIBLE LEAD EFFECT |1

Enter the number of periods to search for a lead variable. This feature, if enabled, will evaluate possible leads for all input series that have names beginning with the string "MOVE". For example sales might arise the week before a holiday. To detect this temporal structure this option has to be enabled and the potential indicator series must have a name like MOVEX1 or MOVEX2.

Line 9

USE DIFFERENCE FACTORS FROM ARIMA IN TF MODEL |1

A number or researchers have found that while differencing factors are important to Transfer Function Identification, they may be counter-productive when included in the actual estimated model. Since this version of AUTOBOX BATCH is rich in model augmentation procedures(step-up.... sufficiency), it may be possible to simplify the initial structure and then evolve via model augmentation procedures to the final model. The user has the choice of including the ARIMA model differences in the initially identified Automatic model. This feature only effects Automatic Transfer Function initial model identification.

Line 10

CONSTRAIN ALL USER CAUSAL COEFFICIENTS IN MODEL |0

This option allows the user to constrain the final model such that all coefficients for user input series or model developed series, such as interventions, will be kept REGARDLESS of level of significance. Thus those coefficients that AUTOBOX BATCH would have considered insignificant and would have replaced with a zero instead become part of the model. This could be particularly interesting to the user who would like to see the affect of his causative input series instead of having them ignored be- cause of their perceived insignificance. This is then a CONSTRAINED REGRESSION option where certain coefficients are a permanent part of the model. Note that this does not constrain the actual values of the parameters.

Line 11

ENABLE MODEL ESTIMATION |1

Estimation and diagnostic checking represent the second phase of the B-J modeling procedure. The estimation option computes the model coefficients and the residual statistics via non-linear least squares.

Line 12

OF GROUPS IN POOLED-CROSS SECT. T/S (IF ANY) | 1

POOLED-CROSS SECTIONAL TIME SERIES

Consider the case where you have n distinct time series (max of 3) and you wish to test the hypothesis that the individual ARIMA models are equal to each other vs. the alternative that at least 1 model differs from the rest. This requires that 1 model be specified for all n and parameter estimation be done locally and compared to a global or generic set of coefficients. A STARTING MODEL MUST EXIST as this will be used. If AUTOMATIC MODELING IS DISABLED and this answer is greater than one (1) the program will: 1. disable all model modification options (sufficiency, necessity etc) 2. expect the time series to be a concatenated series of the n distinct time series and will estimate parameters without using the last set of group i to predict the start of i+1 , where i goes from 1 to 2 (max 3 groups). Hypothesis testing is done by summing the error sum of squares from the n local estimations (done separately) and divide by the total degrees of freedom to obtain a denominator mean square error. The numerator mean square error is the differential error sum of squares (composite estimation less the sum of the locals, divided by the number of groups see JOHNSTON : ECONOMETRIC METHODS 1963 Page 137)

Line 13

SAMPLE SIZE IN EACH OF THE GROUPS |

Enter the number of observations in each of the groups. If you specified n groups in the concatenated series then you must now enter the n values indicating the number in EACH group, in the same sequence the groups were entered into the concatenated series.

Line 14

% CHANGE IN ERROR SUM OF SQUARES |00.1

Parameter estimation is an iterative process that stops when one of three conditions is met. If the relative change in the residual sum of squares is less than the value specified here, then the parameter estimation will stop.

Line 15

% CHANGE IN THE PARAMETER VALUES |00.1

Parameter estimation is an iterative process that stops when one of three conditions is met.. If the relative change in each individual parameter is less than the value specified here, then the parameter estimation will stop.

Line 16

MAXIMUM # OF ITERATIONS |20

Parameter estimation is an iterative process that stops when one of three conditions is met.. If the number of iterations in the estimation process exceeds the value specified here, then the parameter estimation will stop.

Line 17

NECESSITY TEST: | 1

Choose "1" to enable diagnostic checking. One phase of diagnostic checking entails deleting unnecessary parameters from the model. This normally requires you to re-specify the model form, and then to estimate this model. With this option on, the program automatically deletes the non-significant parameters (one at a time) and re-estimates the model. The test for necessity is performed by examining the T-ratios for the individual parameter estimates. Parameters with nonsignificant coefficients will be deleted from the model.

Line 18

CONFIDENCE LEVEL FOR NECESSITY |95.0

If you elected to turn the parameter deletion option on, then you have the option of specifying the confidence level value that will be used to determine the significance of a parameter. For example, 95% indicates that the program should delete all parameters that are not significant at the 95% level.

Line 19

SUFFICIENCY TEST(STOCHASTIC STRUCTURE) | 1

The diagnostic checking phase requires the analyst to make sure that the residuals can not be predicted from themselves (ACF) and in the case of multivariate models the pre-whitened input series (CCF). In the latter case, the test also has to be reversed, i.e. the residuals can not predict the pre-whitened input, otherwise the condition of feedback is identified. The residuals are tested for white noise in much the same way as model identification is performed. If there are patterns in the residual autocorrelations and partial autocorrelations, then the analyst may need to add parameters to the model. One follows the pattern recognition rules described above when adding parameters to the model. A "yes" will request the program to guide these adjustments.

Line 20

CONFIDENCE LEVEL FOR SUFFICIENCY (SS) |95.0

If you elect to augment an estimated model with additional ARIMA structure as evidenced by the sample ACF and PACF of the residuals, you must indicate the significance of a parameter. For example, 95%

Line 21

SUFFICIENCY TEST(DETERMINISTIC STRUCTURE) | 1

Outliers can occur in many ways. They may be the result of a gross error, for example, a recording or transcript error. They may also occur by the effect of some exogenous intervention. These can be described by two different, but related, generating models discussed by Chang and Tiao (1983) and by Tsay (1986).They are termed the innovational outlier (IO) and additive outlier (AO) models. AUTOBOX BATCH uses the AO approach due to estimation considerations. ARIMA modeling may be deficient when the series has been intervened with. This program will test the residuals from the ARIMA model for possible outlier (intervention) variables. We suggest that you modify either your model or your time series for any outlier variables that may be found. The automatic intervention detection option automatically determines the need for intervention variables using the residuals from an estimated model and automatically introduces them into the model.

Line 22

CONFIDENCE LEVEL FOR SUFFICIENCY (DS) |90.0

If you select the outlier detection option, then you must specify the confidence limit to be used for detecting possible outlier variables. For example, .80 indicates that the program should identify all outliers that are significant at the 80% level.

Line 23

MAXIMUM NUMBER OF OUTLIERS TO BE IDENTIFIED | 5

You may elect to limit AUTOBOX BATCH to a certain number of empirically identified outliers. As delivered, the standard product is limited to a maximum of 5 input series in a transfer function thus this integer can't exceed that limit. AFS sells larger versions which allow up to 150 input series. This feature allows the user to control the incorporation of potentially spurious interventions leading to numerical instability.

Line 24

INCLUDE PULSE VARIABLES | 1

Select "1" to include pulse interventions.

Line 25

INCLUDE STEP VARIABLES | 1

Choose "1" to include step interventions.

Line 26

MINIMUM NUMBER OF OBSERVATIONS IN GROUP | 2

The number entered determines how many successive values that are on a different level, before Autobox batch will consider there to be a level shift.

Line 27

INCLUDE SEASONAL PULSE VARIABLES | 1

Choose "1" to include seasonal pulse interventions.

Line 28

INCLUDE LOCAL TRENDS |0

Choose "1" for Autobox batch to identify multiple trends.

Line 29

ENABLE AUTOMATIC FIXUP FOR SEASONAL DUMMIES |0

Choose "1" to enable this option to test for the presence of a SEASONAL DETERMINISTIC VARIABLE which has a zero/one pattern according to the following:

a "1" in the corresponding period and a "0" in other periods

The formal test is outlined in Franses paper in the International Journal Of Forecasting, July 1991, pp 199-208 (see the help for the associated Confidence value indicates that the program should add all parameters that appear to be needed at the 95% level.

Line 30

NUMBER OF LAMBDA VALUES TO EVALUATE in EST |0

Enter the number of values to be included in line 59. If you indicate 3, you must supply three values in the next prompt.

Line 31

LAMBDA VALUES TO EVALUATE |1.0 0 -.5

INITIAL MODEL IDENTIFICATION

The lambda value is the transformation parameter. In other words, the value that you specify here will be the exponent in the power transformation. Each data point in the time series is raised to the power lambda. The acceptable range of lambda values is from 1.0 to -1.0. For example, a lambda of 1.0 indicates that the

original series is to be analyzed, a value of 0.0 indicates that the natural log of the series is to be analyzed, and a lambda of -1.0 indicates that the inverse of the series is to be analyzed. It represents the power transformation that is to the observed series in order to induce variance stationarity. Note however that this should only be applied when the non-constant variance is caused by a correlation between level and variability.

Line 32

DISCRETE CHANGE TEST FOR VARIANCE | 0

The residuals from a model may not have constant variance and consequently the standard estimation may be deficient. One form of non-constant variance is treated by the Box-Cox or lambda transformations. However, a different kind of non-constancy can occur if a series is affected by a period of unusual volatility. Consider the case where an upward trending series has a residual variance of say 10 for the first half and a variance of the residuals of 20 for the second half. It would be totally incorrect to either ignore the change in variance or to use the power transform procedures of Box-Cox. The suggested procedure is to simply identify a model and compute a vector of residuals. By breaking the residuals into consecutive but non-overlapping sections one can perform the standard F test for variance change. The time period with the greatest F value is then a potential point of variance change. There are cases in which the user has an "a priori" knowledge of the weights and wishes to estimate a user-specified model or to automatically build one using these weights or pre-assigned "degrees of believability". For example consider, the actuarial economist who has chronological data where the reading at time period t is based upon "n" samples. Thus a reading with a large "n" is more credible than one with small "n". The user can enter the weights in a disk file Weights.In . These weights will then be used and will be potentially modified if this test is enabled. The resultant weights are stored in Weights.Out, if the I/O option is specified.

Line 33

CONFIDENCE LEVEL FOR VARIANCE TEST |90.0

If you select the variance stability test, then you must specify the confidence limit to be used for detecting possible change points. For example, .95 indicates that the program should identify all time periods that are significant at the 95% level. The interval for comparing variances (Variance Stability) is based on the number of forecasts. If you specify a 3 period forecast, then testing for variances (if enabled) will be done until a group of residuals is less than 3. If the number of forecasts is 0, the program defaults to a minimum of 10 residuals in a group.

Line 34

MINIMUM NUMBER OF RESIDUALS TO POOL | 5

This entry controls the length of the interval for comparing variances . For example if you specify a 10 , then a minimum of 10 residuals will be pooled and compared against the remaining residuals.

Line 35

CONSTANCY OF PARAMETERS: | 0

Choose "1" to enable. Chow suggested a test to assess the statistical significance between two sets of regression coefficients. We have extended this to ARIMA/TF models and furthermore SEARCH for the point of maximum contrast thus identifying the local cluster of homogenous data. This is equivalent to pooled cross-sectional time series where the number of consecutive values in each of the two groups is unknown and to be determined.

Line 36

CONFIDENCE LEVEL FOR CONSTANCY TEST |

If you select the variance stability test, then you must specify the confidence limit to be used for detecting possible change points. For example, .95 indicates that the program should identify all time periods that are significant at the 95% level. The interval for comparing variances (Variance Stability) is based on the number of forecasts. If you specify a 3 period forecast, then testing for variances (if enabled) will be done until a group of residuals is less than 3. If the number of forecasts is 0, the program defaults to a minimum of 10 residuals in a group.

Line 37

STORE MODEL FORM (YES/NO) |1

Choose '1' if you have a model saved and want to reuse it now.

Line 38

DISPLAY MANAGEMENT ANALYSIS | 0

If you want a report that tries to summarize "in english" information about the time series from the model used to fit the data.

Line 39

ENABLE MODEL FORECASTING | 1

The forecasting program generates the forecast values for each time series. This option should be selected upon successful identification of the final Box-Jenkins model form.

Line 40

OF FORECAST VALUES TO COMPUTE | 10

This entry indicates how many forecast values you want the program to compute. The accuracy of the forecasts can be assessed by the resulting errors. Since this entry measures the length of the interesting interval, it is also used to control the interval for comparing variances (Variance Stability). If you specify a 3 period forecast, then testing for variances (if enabled) will be done until a group of residuals is less than 3. If the entry here is 0, the program defaults to a minimum of 10 residuals in a group.

The maximum is the seasonality multiplied by 3. So if the maximum seasonality of 60 is used then the maximum number of forecasts to compute is 180.

Line 41

CONFIDENCE LIMIT FOR THE FORECASTS (%) |80.0

The reliability of a forecast is measured in terms of its uncertainty. This program will compute the individual confidence limit of each forecast, given the information available at the forecast origin. You can specify whatever percent confidence limit you want the program to use.

Line 42

CONVERT THE FORECAST VALUES TO POSITIVE VALUES | 0

If your data arises only in positive values then you might wish to constrain forecasts to the set of positive real numbers. A "YES" will convert the forecasts and confidence limits. All error reports are presented in terms of these rounded forecasts. Note that the aggregated sum is rounded after the aggregation thus the sum of the forecasts may not be equal to the aggregated sum.

Line 43

CONVERT THE FORECAST VALUES TO INTEGERS | 0

If your data arises only in integer form, then it is known as DARMA or a discrete ARIMA problem. One can approximate a DARMA model by estimating as if the data were continuous and then integerizing the forecasts. This is an approximation and the user should be guided by the results. A 'yes' will convert the forecasts and confidence limits to the nearest integer. Thus the forecasts will be rounded off rather than truncated. All error reports are presented in terms of these rounded forecasts. Note that the aggregated sum is rounded after the aggregation thus the sum of the forecasts may not be equal to the aggregate of the forecasts. ARIMA models are also an approximation to a process that is continuous and is sampled at fixed intervals. Again the ARIMA model is an approximation and the user should be guided by the results.

Line 44

CONVERT PULSE AT LAST OBSERVATION TO STEP | 0

Allows the user to apply his knowledge that the last observation is not a pulse but a permanent step that must be considered.

Line 45

CONVERT PULSE TO SEASONAL PULSE (SAVE LAST OBS | 0

You may elect to convert an identified pulse at a particular time period to a SEASONAL PULSE. Consider where insufficient data exists to confirm a SEASONAL PATTERN. This feature allows the user to enforce the rule that all pulses, save a pulse at the last observation will be treated as the first point in a repetitive pattern.

Line 46

DISPLAY IDENTIFICATION INFORMATION | 1

If no detail is required in the initial identification process send a "no". Autocorrelation is a measure of the unconditional dependence that exists between observations in a time series that are separated by a particular time interval, called lag. The value of the autocorrelation lies between +1 and -1. The closer the autocorrelation is to +1 and -1, the more highly correlated are the observations separated by the particular lag being considered. In summary, the autocorrelation measures the unconditional relationship between lags. Partial Autocorrelation is a measure of the conditional dependence that exists between observations in a time series that are separated by a particular time interval. The value of the partial autocorrelation lies between +1 and -1 and is evaluated just like the ACF. In summary, the partial autocorrelation measures the conditional correlation between lags. Cross Correlation is a measure of the dependence that exists between observations in two time series that are separated by a particular time interval, called lag. The value of the cross correlation lies between +1 and -1. The closer the cross correlation is to +1 and -1, the more highly correlated are the observations separated by the particular lag being considered. If the correlation is closer to +1, a positive correlation is indicated; if it is closer to -1, a negative correlation exists. In summary, the cross correlation measures the strength of the relationship between the lags of two time series.

Line 47

DISPLAY ACF TABLE DURING IDENTIFICATION | 1

Choose "1" to see a table of the correlations at the initial identification stage. The table displays rows of correlations and their standard errors.

Line 48

DISPLAY ACF GRAPH DURING IDENTIFICATION | 1

Choose "1" to see a plot of the correlations at the INITIAL identification stage.

Line 49

DISPLAY CCF TABLE DURING IDENTIFICATION | 1

Choose "1" to see the cross-correlations between the prewhitened input and the prewhitened output series. This information is a statistical 'tool' used to identify the form of a transfer model. This option allows you to control whether or not they get reported in a table.

Line 50

DISPLAY CCF GRAPH DURING IDENTIFICATION | 1

Choose "1" to see the cross-correlations between the prewhitened input and the prewhitened output series. This information is a statistical 'tool' used to identify the form of a transfer model. This option allows you to control whether or not they get reported in a plot.

Line 51

DISPLAY PREWHITENING MODEL(S) | 1

Choose "1" to see the model form as a table.

Line 52

DISPLAY PREWHITENING MODEL(S) EQUATION | 1

Choose "1" if you want the program to display the model(s) in the form of an equation.

Line 53

DISPLAY IDENTIFIED MODEL | 1

Choose "1" if you wish to see the model form as a table.

Line 54

DISPLAY IDENTIFIED MODEL EQUATION | 1

Choose "1" if you want the program to display the model(s) in the form of an equation.

Line 55

DISPLAY ESTIMATION INFORMATION | 1

If this is set to "0" then no estimation information will be reported. This means that lines 123-128 would be skipped.

Line 56

DISPLAY ESTIMATED MODEL PARAMETERS | 1

Choose "1" if you wish to see the model form as a table.

Line 57

DISPLAY ESTIMATED MODEL EQUATION | 1

Choose "1" if you want the program to display the model(s) in the form of an equation.

Line 58

DISPLAY UNTRANSFORMED FIT VS ACTUAL TABLE | 1

Choose "1" so that the program displays a chart which shows the fit values, the residual values and the actual values from the estimated model.

Line 59

DISPLAY DIAGNOSTIC CHECKING INFORMATION | 1

Choose "1" to show detail regarding the diagnostic checking process.

Line 60

DISPLAY RESIDUAL ACF TABLE | 1

Choose "1" to see a table of the residual correlations each time that they are computed, the entry here should be a 'yes'. The table displays rows of correlations and their standard errors.

Line 61

DISPLAY RESIDUAL ACF GRAPH | 1

Choose "1" to see a plot of the residual correlations each time that they are computed, the entry here should be a 'yes'.

Line 62

DISPLAY RESIDUAL CCF TABLE | 1

Choose "1" to have Autobox batch display the cross-correlations between the prewhitened input and the residuals from the current model are the statistical 'tool' used to identify the form of a fixup required to the transfer model. This option allows you to control whether or not they get reported in a table.

Line 63

DISPLAY RESIDUAL CCF GRAPH | 1

The cross-correlations between the prewhitened input and the residuals from the current model are the statistical 'tool' used to identify the form of a fixup required to the transfer model. This option allows you to control whether or not they get reported in a plot.

Line 64

DISPLAY NECESSITY TEST RESULTS | 1

Choose "1" to see the necessity test results. ARIMA modeling may be deficient when the model has too many coefficients. It is important to discard or delete unnecessary structure as it inflates forecast variances, among other things.

Line 65

DISPLAY SUFFICIENCY TEST RESULTS | 1

Choose "1" to see the sufficiency test results. ARIMA modeling may be deficient when the MODEL does not have enough structure. The omitted structure can be identified by studying the sample ACF AND PACF of the residuals. In this way we move structure from the residuals to the model.

Line 66

DISPLAY VARIANCE STABILITY TEST RESULTS | 0

Choose "1" to see the variance stability results. ARIMA modeling may be deficient when the series has a non-constant variance. The program will test the residuals from the ARIMA model for possible change points. Essentially interventions are changes in the mean level of the errors while variance stability measures changes in the variance.

Line 67

DISPLAY OUTLIER TEST RESULTS | 1

Choose "1" to enable this function. ARIMA modeling may be deficient when the series has been intervened with. This program will test the residuals from the ARIMA model for possible outlier (intervention) variables. We suggest that you modify either your model or your time series for any outlier variables that may be found. If you have enabled automatic fixup FOR outliers, in the choose analysis options section then these modifications will be done for you automatically. A "yes" shows the details of this process.

Line 68

DISPLAY CONSTANCY TEST | 0

Choose "1" to see the constancy test results. You get a table showing the observations and those values with a significant change in the reliability of the model parameters.

Line 69

DISPLAY FORECASTING INFORMATION | 1

If this is set to "0" then no forecasting information will be reported. This means that lines 143-150 would be skipped.

Line 70

DISPLAY FORECAST MODEL PARAMETERS | 1

Choose "1" to see the model form as a table.

Line 71

DISPLAY FORECAST MODEL EQUATION | 1

Choose "1" if you want the program to display the model(s) in the form of an equation.

Line 72

DISPLAY TABLE OF TRANSFORMED FORECAST VALUES | 1

Choose "1" to see the forecast in transformed units. Your forecasting model may contain a transformation parameter (lambda). If so, then the program generates forecast values for both the original data and the transformed data.

Line 73

DISPLAY TABLE OF FORECAST VALUES | 1

Choose "1" to see the forecasts with their confidence bounds. Your forecasting model may contain a transformation parameter (lambda). If so, then the program generates forecast values for both the original data and the transformed data.

DISCUSSION OF FFLITE

Assuming that the series 0438.asc had been analyzed, the compressed model saved in 0438.abl would look as follows:

```
     147/  2 745 13  7 40  7  2    1.000
HIGH            0 13  0
MOVE_PRICE      0  17 4
0438            0
 1
 0          0.0084611576
 1
 4        -0.4498213812
 1
 0          0.0000000000
*****BASELINE****
        3.1871258281        0.7146756892
        3.0262497639        0.7146756892
        3.4692965832        0.7146756892
        3.3413501075        0.7146756892
        3.4936723981        0.7146756892
        3.1557451419        0.7146756892
        3.1244239210        0.7146756892
        3.1871258281        0.7146756892
        4.0408596480        0.7146756892
        3.4692965832        0.7146756892
        3.3413501075        0.7146756892
        4.7164280232        0.7146756892
        3.1557451419        0.7146756892
```

What we know after reading **line 1**

```
     147/  2 745 13  7 40  7  2    1.000
```

The date of the LAST OBSERVATION is 147/2 based upon 745 historical values.

A forecast of 13 periods was made with seasonality = 7. The first data point was at 40/7 and there are 2 DYNAMIC VARIABLES (DY) in the model and the transformation parameter = 1.000 meaning no transformation is required

PLEASE NOTE: YOU SHOULD ARCHIVE THE SUM OF THE ORIGINAL NUMBER OF HISTORICAL VALUES AND THE ORIGINAL NUMBER OF FORECAST PERIODS(758=745+13) AS A CONTROL NUMBER FOR UPDATING THE FORECASTS. IN UPDATING, THE SUM OF ANY NEW NUMBER OF HISTORICAL VALUES AND/OR NEW NUMBER OF FORECAST PERIODS CANNOT EXCEED THE CONTROL NUMBER. IF YOU DESIRED TO EXCEED THAT CONTROL, YOU WOULD HAVE TO PROCESS THE SERIES WITH THE UPDATED DATA BY GOING THROUGH THE ENTIRE MODELING PROCESS IN TOTAL.

> The next number of DY lines indicate the series name(position 1-22), the number of historical values required(position 25-28), the number of future values required(position 29-32); and the number of required lead values(position 33-36).

> The next line indicates the output series name(position 1-22) and the number of historical values required(position 25-28)

The next lines indicate, for each series, a) the number of coefficients, b) the backorder power and the related coefficient. **These are not to be changed**

```
        1
        0          0.0084611576
        1
        4         -0.4498213812
        1
        0          0.0000000000
```

The *****BASELINE**** data reflects the original baseline forecasts in the left column and their related degree of accuracy in the right column. Baseline forecasts were generated in AUTOBOX using all the factors of the full model except the Dynamic series. **These values cannot be changed.**

RULES TO DETERMINE # OF HISTORICAL VALUES NEEDED BY THE MODEL

We will read the next DY+1 records which in this case is 3.and examine the 25-28th position to determine the number of historical values required for each of the DY+1 series.

E.G. the next 3 records or lines are:

```
        HIGH              0   13   0
        MOVE_PRICE        0   17   4
        0438              0
```

0 + 0 + 0 gives us 0 and we have no need for historical values.

If, for example, the 3 records were:

```
HIGH           5  13  0
MOVE_PRICE     3  17  4
0438           7
```

then we would need 5+3+7 or 15 values in the 0438.his file, the supporting file containing the required historical values.

If historical values were needed, then the file would be saved with the number of historical values required for each of the DYNAMIC SERIES in the order indicated, followed by the number of historical values required for the OUTPUT SERIES. The values for each series would be entered from oldest to latest (that is: most recent data).

RULES TO DETERMINE # OF FUTURE VALUES NEEDED BY THE MODEL

The number of future values for each dynamic series is the number of forecasts in line 1, position 21 – 24 plus the number of leads indicated in position 33-36 in the respective DY lines. For HIGH there are 0 leads, so the number of future values in the 29-32th position will be 13 (13 + 0). For MOVE_PRICE the number of future values in the 29-32th position will be 17 (13 + 4).

Thus 13+17 gives 30 and we need 30 values for the 0413.fut file, the companion file containing the required future values. The first 13 values reflect "guesstimates" or planned values for the first DYNAMIC SERIES named "HIGH".

The next 17 values reflect "guesstimates" or planned values for the second DYNAMIC SERIES named "MOVE_PRICE".

Note, there are never any future values or leads indicated for the output series

Procedures:

TO OBTAIN REVISED FORECASTS AS NEW SALES DATA BECOMES AVAILABLE.

Consider that 3 periods have gone by and we have just observed OBSERVATION 748 at time period 147/5; and we wish to forecast out the next 9 days.

The new historical values are as follows:

#	date	sales	HIGH	MOVE_PRICE
746	147/3	3	5	6.10
747	147/4	2	7	6.05
748	147/5	1	3	6.15

We have to modify the three pieces of information ..

Model

Historical values

Future values

<u>Model</u>

> We would change the first record to the following reflecting the NEW YEAR OF THE LAST OBSERVATION(position 7-10), NEW PERIOD OF THE LAST OBSERVATION(position 12-15) THE NEW TOTAL OF HISTORICAL VALUES(position 16-20) and the NEW NUMBER OF FORECASTS(position 21-24):

147/ 5 748 9 7 40 7 2

Please note: as indicated in the Explanations above, the sum of any new number of historical values and/or new number of forecast periods cannot exceed the sum of the original number of historical values and the original number of forecast periods

The next set of records up to and including "BASELINE" would be unchanged except for the new future values required. Again as explained in the Rules above, the number of future values for each dynamic series is the number of forecasts in line 1, position 21 – 24 plus the number of leads indicated in position 33-36 in the respective DY lines. For HIGH there are 0 leads so the number of future values in the 29-32th position will be 9 (9 + 0). For MOVE_PRICE the number of future values in the 29-32th position will be 13 (9 + 4).

.

```
HIGH              0   9   0
MOVE_PRICE        0   13  4
0438              0
  1
  0        0.0084611576
  1
  4       -0.4498213812
  1
  0        0.0000000000
*****BASELINE****
```

The next 9 records or lines would be taken from the original baseline data starting with the 4th line, since we now have real data for the first three periods.

```
3.3413501075      0.7146756892
3.4936723981      0.7146756892
3.1557451419      0.7146756892
3.1244239210      0.7146756892
3.1871258281      0.7146756892
4.0408596480      0.7146756892
3.4692965832      0.7146756892
3.3413501075      0.7146756892
4.7164280232      0.7146756892
```

Historical Values

If historical values were needed, as indicated in position 33-36 in the respective DY lines and the OUTPUT SERIES line, then the array would be loaded using the number of historical values required for each of the DYNAMIC SERIES in the order indicated, followed by the number of historical values required for the OUTPUT SERIES; however they would now incorporate the latest historical values. The values for each series would be entered from oldest to latest ..most recent.

In our example case there are no required historical so nothing would be added to the 0413.his file.

Future Values

Examining the required number of future values (position 29-32) tells us we need provide 22 future values, 9 for the first DYNAMIC SERIES named HIGH and 13 for the second DYNAMIC SERIES.

Autobox

Reference Guide
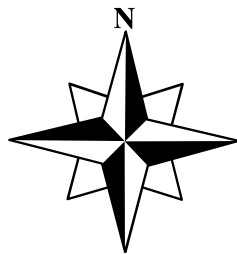
N

Contents

1

Overview

1. General Remarks

Autobox from Automatic Forecasting Systems (AFS) is an expert system which can be used to model and forecast both univariate and multivariate time series. Its methodology is based on the Box-Jenkin's ground breaking body of work. Outliers or intervention variables can be detected and incorporated into both the non-causal (univariate) and the causal (transfer functions) models. In addition, variance constancy and parameter constancy can be tested and remedies developed. The Autobox system has been under development (see references) since 1976 by AFS.

Autobox is attractive to both the expert and the non-expert. The expert can apply his knowledge and control the flow of modeling and estimation. The non-expert can use it as a "statistician-in-the-box" or as a "productivity aid" as it delivers a cost-effective solution. Autobox offers sophisticated analytical tools which can be recommended to teachers, academic researchers and forecasters who usually analyze a handful of time series with great care. Autobox offers a total automation of the modeling and forecasting tasks. Therefore, it can also be recommended to practitioners who must process large amounts of data automatically.

AFS was the first company to automate the Box-Jenkins model building process. Our approach is to program the model identification, estimation and diagnostic feedback loop as originally described by Box and Jenkins. This is implemented for both ARIMA (univariate) modeling and transfer function (multivariate or regression) modeling. What this means is that the user from novice to expert can feed Autobox any number of series and the program's powerful modeling heuristic can do the work for you. This option is implemented in such a way that it can be turned on at *any* stage of the modeling process. There is complete control over the statistical sensitivities for the inclusion/exclusion of model parameters and structures. These features allow the user complete control over the modeling process. The user can let Autobox do as much or as little of the model building process as you or the complexity of the problem dictates.
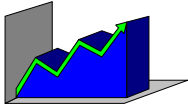
2. Modeling Overview

Autobox is a forecasting engine built on the central modeling steps of the Box-Jenkins paradigm. This core is extended by several useful modules: intervention detection, a simulation option and numerous others, all of which expedite the forecasting practitioner's tasks.

Autobox allows a single endogenous equation incorporating either user specified candidate causal series or empirically identified dummy series. The set of user specified candidate causal series can be either stochastic or deterministic (dummy) in form. During the search for the most appropriate model form and the optimal set of parameters the program can either be:

1. Purely empirical or

2. A starting model could be used.

A final model may require one or more of the following structures:

1. Power transformation like Log, Square Root, Reciprocal, etc.

2. Variance stabilization due to deterministic changes in the background error variance.

3. Data segmentation or splitting as evidenced by statistically significant changes in either model form or parameters.

En route to its tour de force, Autobox will evaluate numerous possible models/parameters that have been suggested by the data itself. In practice, a realistic limit is set on the maximum number of model form iterations. The exact specifics of each tentative model are not pre-set thus the power of Autobox emerges. The kind and form of the tentative models may never before been tried. Each data set speaks for itself and suggests the iterative process. The final model could be as simple as:

1. A simple trend model or a simple ordinary least squares model.

2. An exponential smoothing model.

3. A simple weighted average where the weights are either equal or unequal.

4. A Cochrane-Orcutt or ordinary least squares with a first order AR fixup.

5. A simple ordinary least squares model in differences containing some needed lags.

---

6. A spline-like set of local trends superimposed with an arbitrary ARIMA model with possible pulses.

The number of possible final models that Autobox could find is infinite and only discoverable via a true expert system.

A final model may require one or more of the following seasonal structures:

1. Seasonal ARIMA structure where the prediction depends on some previous reading S periods ago.

2. Seasonal structure via a **complete** set of seasonal dummy's reflecting a fixed response based upon the particular period.

3. Seasonal structure via a **partial** set of seasonal dummy's reflecting a fixed response based upon the particular period.

The final model will satisfy both:

1. Necessity tests that guarantee the estimated coefficients are statistically significant.

2. Sufficiency tests that guarantee that the error process is:

-unpredictable on itself

-not predictable from the set of causals or their lags (leads).

-has a constant mean of zero.

The final model will contain one or more of the following structures:

1. CAUSAL with correct lead/lag specification.

2. MEMORY with correct "autoregressive model components".

3. DUMMY with correct pulses, seasonal pulses, level shifts or spline time trends.

3. Statistical Features

Autobox considers the same model class, namely the seasonal autoregressive integrated moving average model with dynamic regressors (transfer components), abbreviated as SARIMAX(p,d,q)(P,D,Q)s[X]. The identification, estimation and outlier diagnosis issues are attacked in a rather elegant way via maximum likelihood estimation.

The user is responsible for selecting the dependent and independent series, and the time range so that Autobox will produce automatic forecasts. The way this is done is controlled by 155 switches specified by the user in various sub-menus.

Switches control the extent of automatic identification steps such as the use of Box-Cox transformation for variance stabilization, setting the confidence levels, the way how to handle differences to achieve stationarity and so on.

The results of the different modeling alternatives can be reviewed by setting all 45 switches governing "output report options" to a "YES". This yields huge amounts of output, but reveals at least some part of the inherent modeling wisdom of the program. Autobox uses several advanced and clever techniques to identify suitable forecast models, including necessity and sufficiency tests for the parameters, a stability check of the ARMA polynomials (stationarity and invertibility). In keeping with the latest methodologies, we have applied a recently developed test by Franses to discriminate between seasonal unit roots and seasonal dummy model components.

Outlier handling in Autobox recognizes several other types of outliers which **<u>cannot</u>** be handled by the other programs. One of these is the seasonal pulse, which can be very useful to model seasonal patterns which occur only in a few months (like Christmas effects in December). Another is the step or level shift variable which characterizes a permanent change in the level of the series. A recent advancement is the incorporation of trend detection where the model can

have a number of trends each with their own slope.

Autobox can detect dynamic patterns such as transient changes in the data (i.e. level shifts in the data can be detected in both systems as additive outliers in the first differences). It is also able to detect deterministic trend changes which are increasingly used by econometricians.

Different types of exponential smoothing schemes also can be estimated. Moreover, the parameters of the smoothing procedures are automatically optimized. More significantly for ARIMA users, Autobox extends its automatic modeling to causal modeling.

It offers two causal model identification options. The first technique is the well known prewhitening method developed by Box and Jenkins. As an alternative, the user can select the common filter/least squares technique which is useful when the input causal series are cross-correlated.

Autobox uses the theoretical covariances for alternative candidates as the scheme for both ARIMA and TRANSFER FUNCTION identification. This rather sophisticated pattern recognition selects the numerator/denominator polynomial orders for the rational transfer function weights.

Autobox permits the user to easily evaluate the forecast accuracy of the identified model.

First, the user can withhold some observations for forecast evaluation using an out-of-sample analysis. Second, different forecast evaluation measures including the mean error (bias) and mean absolute percentage error (MAPE) can be computed for different forecast horizons without any additional programming.

Autobox comes with a complete set of identification and modeling tools for use in the BJ framework. This means that you have the ability to transform or prewhiten the chosen series for identification purposes. Autobox handles both ARIMA (univariate) modeling and Transfer Function (multivariate) modeling allowing for the inclusion of interventions (see below for more information). Tests for interventions, need for transformations, need to add or delete model parameters are all available. Autocorrelation, partial autocorrelation and cross-correlation functions and their respective tests of significance are calculated as needed. Model fit statistics, including R², SSE, variance of errors, adjusted variance of errors all reported. Information criteria statistics for alternate model identification approaches are provided.

Intervention Detection

One of the most powerful features of Autobox is the inclusion of Automatic Intervention detection capabilities in both ARIMA and Transfer Function models. Almost all forecasting packages allow for interventions to be included in a regression model. What these packages don't tell you is how sensitive **all** forecasting methodologies are to the impact of interventions or missing variables. These packages don't tell you if your series may be influenced by missing variables or changes that are outside the current model.

If a data series is impacted by changes in the underlying process at discrete points in time, both ARIMA models and Transfer Function models will produce poor results. For example, a competitors price change changes the level of demand for your product. Without a variable to account for this change your forecast model might perform poorly. Autobox implements ground breaking techniques which quickly and accurately identify potential interventions (level shifts, season pulses, single point outliers and changes in the variance of the series). These variables can then be included in your model *at your discretion*. The result is more robust models and greater forecast accuracy.

Graphical Analysis Tools

Autobox has a set of graphing tools that help present complex statistical information in a way that is easy and clear at every stage of the forecasting process. For example, graphs of autocorrelation, partial-autocorrelation and cross-correlation functions are all available. Even more incredibly these can be compared to theoretical values for various models forms.

Forecasting and Diagnostics

All forecast packages allow for you to produce forecasts using the models you have constructed. Autobox presents the critical information you need to determine if those forecasts are acceptable. Autobox has options that allow you to analyze the stability and forecasting ability of your forecast model. This is achieved through a series of ex-post forecast analyses. You can *automatically* withhold any number of observations, re-estimate the model form and forecast. Observations are then added back one at a time and the model is re-estimated and reforecast. Forecast accuracy statistics, including Mean Absolute Percent Error (MAPE) and Bias, are calculated at each forecast end point. Thus the stability of the model and its ability to forecast from various end points can be analyzed. Finally, you can optionally allow Autobox to actually re-identify the model form at each level of withheld data to see if the *model form* is unduly influenced by recent observations.

---
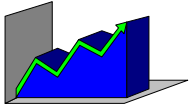
4. Conclusions

Academic researchers who usually analyze a handful of time series with great care and in great detail will be able to accomplish their tasks readily. For this group, the expert system component should be part of a data analysis support tool as it can be a "productivity aid". Quite possibly, they might even find it a replacement (at times) for some of the repetitive tasks and allow them to focus on the "difficult series".

Due to the high integration of the different time series analysis tools into an integrated framework, Autobox can be strongly recommended to practitioners who must handle many data sets. Practitioners with only little knowledge about Box-Jenkins models can use the program as a pure black-box tool, assuming some knowledgeable user with experience in forecasting methods has configured the system to match the specific features of the data and the specific needs of the end user.

Without the support of an experienced forecaster, a novice may unwittingly run into pitfalls. This is probably true for any sophisticated forecasting program not just this one.
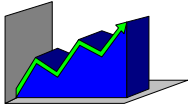
2

Conceptual Overview

Autoregressive Integrated Moving Average (ARIMA) is a process designed to identify a weighted moving-average model specifically tailored to the individual dataset by using time series data to identify a suitable model. It is a rear-window approach that does use user-specified helping variables. It uses correlations within the history to identify patterns that can be statistically tested and then used to forecast. Often we are limited to using only the history and no causals whereas the general class of Box-Jenkins models can efficiently incorporate causal/exogenous variables (Transfer Functions). Thus ARIMA should be more accurately called Univariate Box-Jenkins rather than Box-Jenkins.

The modeling process includes an identification phase. This tutorial will introduce concepts and the steps to identify the model, estimate the model, and perform diagnostic checking to revise the model. We will also list the assumptions and how to incorporate remedies when faced with potential violations.

The Background

Our understanding of ARIMA has grown since it was introduced in (1). Properly formed ARIMA models are a general class that includes all well-known models save some state space and multiplicative Holt-Winters models. While classical ARIMA modeling attempted to capture stochastic structure in the data little was done about incorporating deterministic structure other than a possible constant or identifying change points in parameters or error variance.

In the Exceptions Section we will highlight procedures relevant to suggested augmentation strategies as they were not part of the original ARIMA approach suggested in (1) but are now standard. This step is often ignored as it is necessary that the mean of the residuals is invariant over time and that the variance of the final model errors are constant.
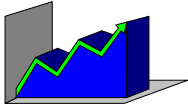
The idea of modeling is to characterize the pattern in the data and the goal is to identify an underlying model that is generating that pattern. The model that you will build should match the history which can then be extrapolated into the future. The actual minus the fitted values are called the residuals. The residuals should be random around zero (i.e. Gaussian) signifying that the pattern has been extracted and captured in the model. In primary school we learned how to determine the general term of both an arithmetic and geometric series, For example the series 1,3,5,7,9,11 would be easily characterized. A series 1,9,1,9,1,9 can also be easily characterized as a process. A series like 1,9,1,9,1,9,5,9,1,9 containing a pulse requires a more sophisticated characterizer as there is deterministic effect at period 7. In general ARIMA modeling is the search for the general term for a given data series.

ARIMA (p, d, q)

Order
of AR

Order
of MA

Order of
"differencing"
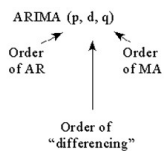
- For example, a AR model for monthly data may contain information from lag 12, lag 24, etc.
    - i.e. $Y_t = A_1 Y_{t-12} + A_2 Y_{t-24} + a_t$
    - This is referred to as an ARIMA(0,0,0)x(2,0,0)$_{12}$ model
    -
- General form is ARIMA(p,d,q)x(ps,ds,qs)$_s$

While seasonal autoregressive structure might be intuitively appealing, early versions of ARIMA premised no seasonal fixed effects such as a February effect or a June effect. Modern ARIMA modeling can suggest the nature of these omitted fixed effects (pulses, seasonal pulses, level/step shifts, local time trends).

The Tools
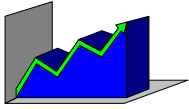
The ARIMA process uses regression/correlation statistics to identify the stochastic patterns in the data. Regressions are run to find correlations based on different lags in the data. The correlation between successive months would be the lag 1 correlation or in ARIMA terms, the ACF of lag 1. Is this month related to one year ago at this time would be the lag 12 correlation or in ARIMA terms, the ACF of lag 12. By studying the autocorrelations in the history, we can determine if there are any relationships and then take action by adding parameters to the model to account for that relationship. The different autocorrelations for the different lags are arranged together in what is known as a correlogram and are often presented using a plot. They are sometimes presented as a bar chart. We present it as a line chart showing 95% confidence limits around 0.0 . The autocorrelation is referred to as the autocorrelation function (ACF).

- The key statistic in time series analysis is the autocorrelation coefficient (the correlation of the time series with itself, lagged 1, 2, or more periods.)
- Recall the autocorrelation formula:

$$r_k = \frac{\sum_{t=k+1}^{n}(y_t - \bar{y})(y_{t-k} - \bar{y})}{\sum_{t=1}^{n}(y_t - \bar{y})^2}$$

Let's consider the Partial Autocorrelation Function (PACF). The PACF of lag 12 for example is a regression using a lag of 12, but also uses all of the lags from 1 to 11 as well, hence the name partial. The ACF could have been called the unconditional regression and the PACF the conditional regression as it measures the conditional impact of adding one more lag. Given this, you should then be able to realize that the PACF of lag 1 is identical to ACF of lag 1. The PACF is ordinarily only needed when performing initial model identification.

- Partial autocorrelations measures the degree of association between y_t and y_{t-k}, when the effects of other time lags 1, 2, 3, ..., k-1 are removed.
- The partial autocorrelation coefficient of order k is evaluated by regressing y_t against $y_{t-1},...y_{t-k}$:
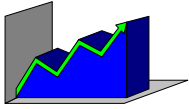
$$y_t = b_0 + b_1 y_{t-1} + b_2 y_{t-2} + \cdots + b_k y_{t-k}$$

Now that we have explained the ACF and the PACF, let's discuss the components of ARIMA. There are three pieces to the model. The "I" means Integrated, but for simplicity sake it really means that you took differencing on the Y variable during the modeling process. The "AR" means that you have a model parameter that explicitly uses the history of the series. The "MA" means that you have a model parameter that explicitly uses the previous forecast errors. Not all models have all parts of the ARIMA model. All models can be re-expressed as pure AR models or pure MA models. The reason we attempt to mix and match has to do with attempting to use as few parameters as possible. For example the somewhat curious model of single exponential smoothing uses 1 coefficient (an MA (1) coefficient) in lieu of an infinite set of AR weights. This model is curious because it expressly restricts the sample space of the so-called smoothing coefficient to ½ the possible space.

Identifying the order of differencing starts with the following initial assumptions, which are ultimately verified:

1) The sequence of errors (a's) are assumed to have a constant mean of 0 and a constant variance for all sub-intervals of time. Note that the inclusion of a constant in the model guarantees that the mean of all the a?s will be 0, but it doesn't guarantee this for each and every sub-interval. Power transformations should never be based upon the properties of the original series but based on the distribution of the estimated a's as that is where the statistical requirements are in effect.
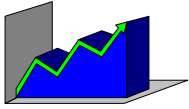
2) The sequence of errors (a's) are assumed to be normally distributed where the a's are independent of each other. Finally the model parameters and error variance are assumed to be fixed over all sub-intervals.

We study the ACF and PACF and identify an initial model. If this initial model is a good model then the residuals will be free of structure and we are done. If not we identify that structure and add that structure to the current model until a subsequent set of residuals is free of structure. This is akin to what happens when your eye doctor creates a new prescription/model for you. One could consider this iterative approach as moving structure in the errors to the model until there is no structure to relocate.

Following are some simplified guidelines to follow when identifying an appropriate ARIMA model with the following assumptions:

* Guideline 1: If the series has a large number of positive autocorrelations then differencing should be introduced. The order of the differencing is suggested by the significant spikes in the PACF based upon the standard deviation of the differenced series.

* Guideline 2: Include a constant if your model has no differencing; include a constant elsewhere if it is statistically significant.

* Guideline 3: Domination of the ACF over the PACF suggests an AR model while the reverse suggests an MA model. The order of the model is suggested by the number of significant values in the subordinate.

* Guideline 4: Parsimony: Keep the model as simple as you can but not too simple as overpopulation often leads to redundant structure.

* Guideline 5: Evaluate the statistical properties of the residual (At) series and identify the additional structure (step-forward) required. Note that the ACF is the ratio of the covariance to the instantaneous variance which means that if there are violations in the At's such as pulse outliers, level shifts , seasonal pulses and/or local time trends this test can be seriously biased downwards leading to the conclusion that there is no additional stochastic structure required. The variance is over estimated due to the pulses/level
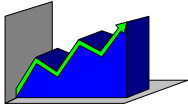
shifts/local time trends thus the ACF is understated. This has been referred to as "The Alice in Wonderland Effect".

* Guideline 6: Reduce the model via step-down procedures to end up with a minimally sufficient model that has effectively deconstructed the original series to signal and noise. Over-differencing leads to unnecessary MA structure while under-differencing leads to overly complicated AR structure.

Guidelines for model form

Process	MA(q)	AR(p)	ARMA(p,q)
Autocorrelation function(ACF)	Cuts off	Doesn't cut off	Doesn't cut off
Partial Autocorrelation function(PACF)	Doesn't cut off	Cuts off	Doesn't cut off

One could consider this a custom-fitting exercise, being cautious not to over fit or under fit. The key question is after estimation, is the current model sufficient? If not we identify the remedial action required via the ACF and PACF of the current (tentative) model's residuals and incorporate that structure into the model. Note that only statistically significant coefficients are used in the final model. If the series did not require differencing then you must have a constant in the model. If the model includes differencing then the constant is optional. In either case we still have to test that our assumptions regarding the errors are correct or at least can't be rejected.
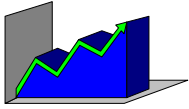
Example 1) We have 36 annual values.

Actuals - WW01

The ACF and the PACF suggest an AR(1) model (1,0,0)(0,0,0).

ACF Originals - WW01

ACF

Lags 1 to 12

● 95% Confidence Level

● ACF
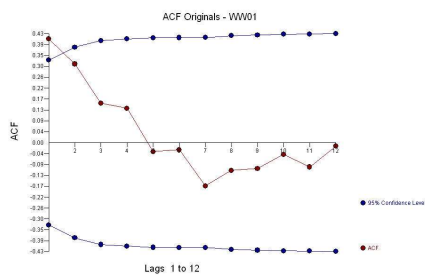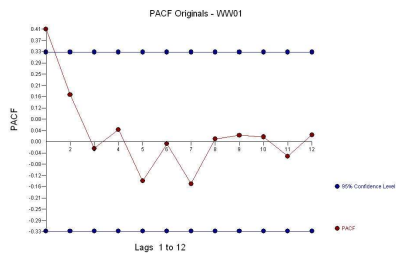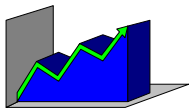
Leading to an estimated model (1,0,0)(0,0,0).

$$Y(T) = .60123 + [(1- .454B** 1)]**-1 [A(T)]$$

With the following residual plot, suggesting some "unusual values".
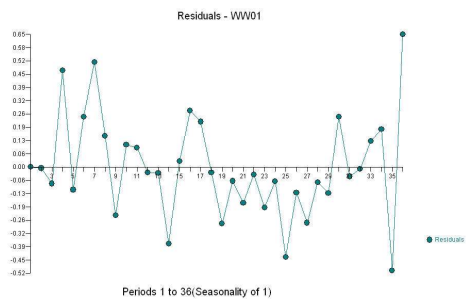
The ACF and PACF of the residuals suggests no stochastic structure as the anomalies effectively downward bias the results.

ACF Residuals - WW01
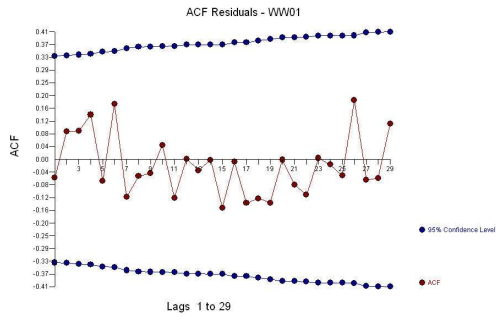
Example 2 has 36 monthly observations:

Actuals - H7OPS

With ACF and PACF :

ACF Originals - H7OPS
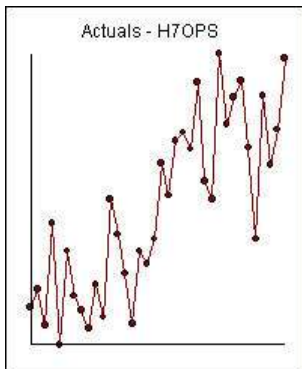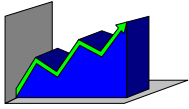
PACF Originals - H7OPS
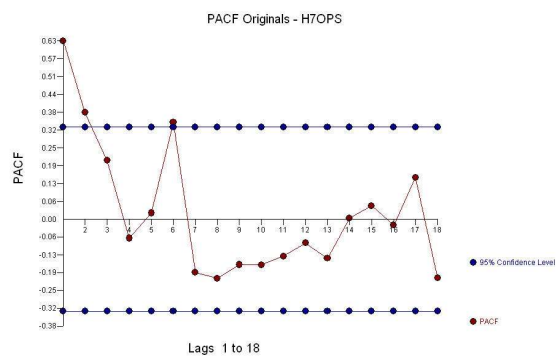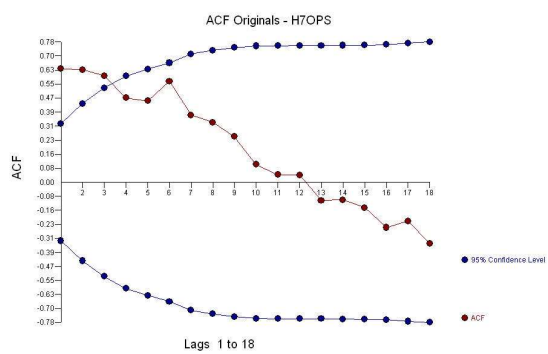
Leading to an estimated model: AR(2) (2,0,0)(0,0,0)

```
Y(T) =   381.96      +     [(1-  .378B** 1-  .462B** 2)]**-1  [A(T)]
```

And with ACF of the residuals:
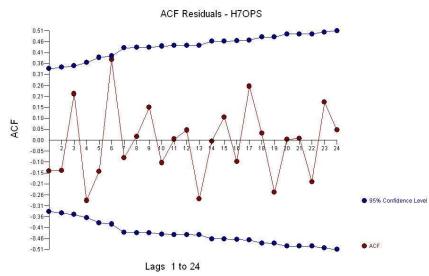
With the following residual plot:
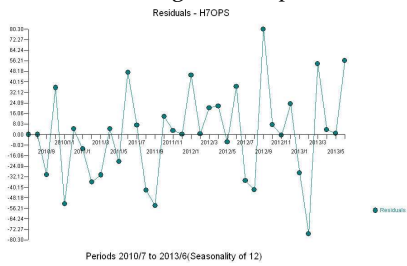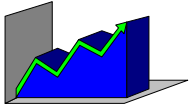
Example 3 has 40 annual values
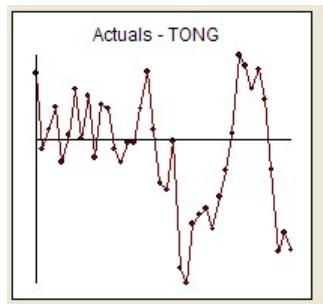
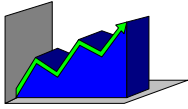The ACF and PACF of the original series are;
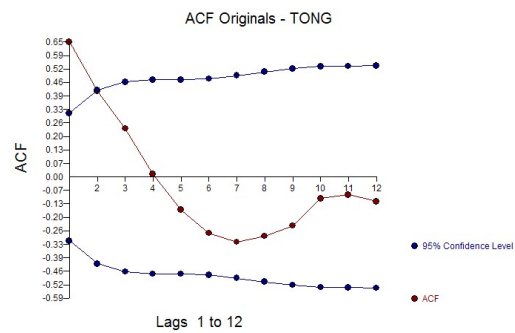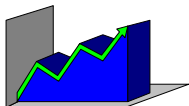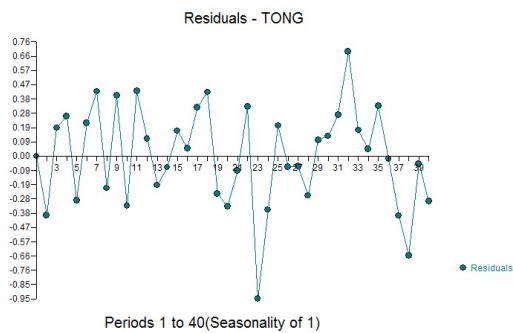
PACF Originals - TONG
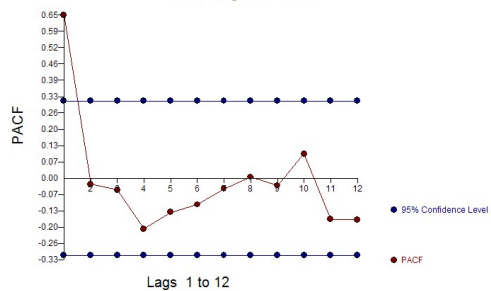
Suggesting a model (1,0,0,(0,0,0)

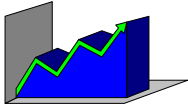$$Y(T) = -.19797 \qquad + \qquad [(1- .706B** 1)]**-1 \ [A(T)]$$

With a residual ACF of

ACF Residuals - TONG

And residual plot

Residuals - TONG

Periods 1 to 40(Seasonality of 1)

Suggesting a change in distribution of residuals during the second half.

Exceptions: If a tentative model exhibits errors that have variance change this can be remedied in a number of ways ; 1) Identify the need to validate that the errors have constant mean via Intervention Detection (2,3) yielding pulse, seasonal pulse/level shift/ /local time trends 2) confirming that the parameters of the model are constant over time 3) confirming that the error variance has had no deterministic change points or stochastic change points. We will focus here on category 1. The tool to identify omitted deterministic structure is fully explained in references 2 and 3. The idea Step 1)is to use the model to generate residuals and then Step 2)identify the intervention variable needed following the procedure defined in reference 3. Step 3). Re-estimate the residuals incorporating the effect into the model and then go back to Step 1 until no additional interventions are found.

We now revisit Example 1 and augment the model with pulse indicators providing a more robust estimate of the ARIMA coefficients.

```
Y(T) = .75857                              WW01
      +[X1(T)][(+ .280)]             :PULSE          7
      +[X2(T)][(+ .441)]             :PULSE          4
      +[X3(T)][(- .746)]             :PULSE         35
      +[X4(T)][(- .254)]             :PULSE         25
      +[X5(T)][(- .302)]             :PULSE         14
      +[X6(T)][(- .300)]             :PULSE          9
      +[X7(T)][(- .198)]             :PULSE         19
      +[X8(T)][(+ .246)]             :PULSE         30
      +     [(1- .859B** 1)]**-1 [A(T)]
```
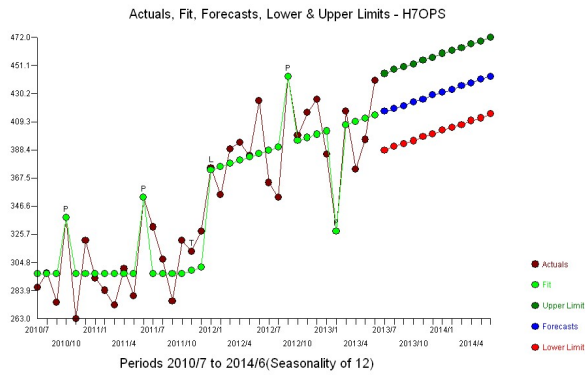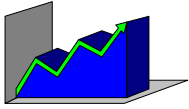
Example 2 is a series that is better modeled with a step/level shift.

The plot of the residuals suggests a mean shift. Empowering Intervention Detection leads to an augmented model incorporating a level shift and a local time trend with and 4 pulses and a level shift. This model is as follows:

```
Y(T) = 296.30                           H7OPS
      +[X1(T)][(+ 2.4008)]   :TIME TREND      17
      +[X2(T)][(- 76.6519)]  :PULSE           32
      +[X3(T)][(+ 69.9367)]  :LEVEL SHIFT      19
      +[X4(T)][(+ 56.6959)]  :PULSE           12
      +[X5(T)][(+ 50.3521)]  :PULSE           27
      +[X6(T)][(+ 41.6959)]  :PULSE            4
      +        [A(T)]
```

Actuals, Fit, Forecasts, Lower & Upper Limits - H7OPS

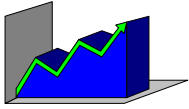Periods 2010/7 to 2014/6(Seasonality of 12)

Example 3

When the parameters were tested for constancy. over time using the Chow Test (5) , a significant difference was detected at period 21.

.....

```
        F TEST TO VERIFY CONSTANCY OF PARAMETERS

  CANDIDATE BREAKPOINT        F VALUE          P VALUE

        18       18           6.0905         .0054861917
        19       19           5.6979         .0073431038
        20       20           6.2987         .0047097652
        21       21           5.5804         .0080202803*
```

As period 1-20 was

```
        MODEL COMPONENT        LAG     COEFF     STANDARD     P        T
  #                            (BOP)             ERROR      VALUE    VALUE

   1CONSTANT                          .963E-01   .587E-01   .1183    1.64
   2Autoregressive-Factor #  1    1  -.122       .229       .6021    -.53
```
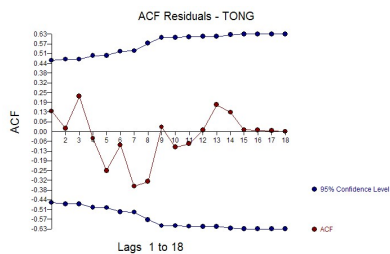
period 21-40

```
        MODEL COMPONENT        LAG     COEFF     STANDARD     P        T
  #                            (BOP)             ERROR      VALUE    VALUE

   1CONSTANT                         -.668E-01   .805E-01   .4175    -.83
   2Autoregressive-Factor #  1    1   .861       .128       .0000    6.72
```

A final model using the last 20 values was

```
[(1-B**1)]Y(T) = +[X1(T)][(1-B**1)][(-  .886)]        :PULSE        23
                 +[X2(T)][(1-B**1)][(+  .286)]        :PULSE        32
                 +[X3(T)][(1-B**1)][(-  .345)]        :PULSE        38
                 +[X4(T)][(1-B**1)][(-  .715)]        :PULSE        24
                 +     [(1-  .453B** 1)]**-1  [A(T)]
```
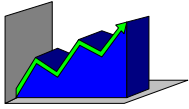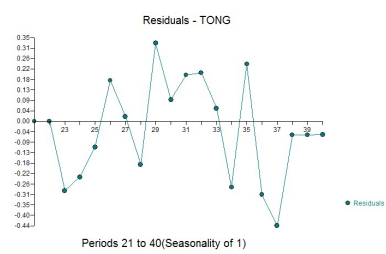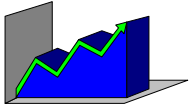
With residual ACF of

ACF Residuals - TONG

And residual plot of

Residuals - TONG

Periods 21 to 40(Seasonality of 1)

3
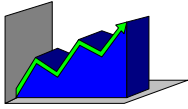
References

Anderson, O.D. (1975). <u>Time Series Analysis and Forecasting</u>. London: Butterworth & Co.

Bell, W. (1983). "A Computer Program for Detecting Outliers in
Time Series," in <u>American Statistical Association 1983 Proceedings of the Business Economic Statistics Section</u>, Toronto, pp. 624-639.

Box, G.E.P., and Cox, D.R. (1964). "An Analysis of Transformations (with discussion)", <u>Journal of the Royal Statistical Society</u>, B26, pp. 212-252.

Box, G.E.P., and Jenkins, G.M. (1976). <u>Time Series Analysis: Forecasting and Control</u>, 2nd ed. San Francisco: Holden Day.
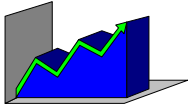
Box, G.E.P., and Tiao, G. (1975). "Intervention Analysis with Applications to Economic and Environmental Problems," <u>Journal of the American Statistical Association</u>, Vol 70,70-79

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.

Downing, D.J., and McLaughlin, S.B. (1986). "Detecting Shifts in Radical Growth Rates by the Use of Intervention Detection," Engineering Physics and Mathematics Division, Oak Ridge National Laboratory, Oak Ridge.

Franses, P., (1991) "Seasonality, non-stationarity and forecasting of monthly time series", International Journal of Forecasting", (7), pp 192-208.

Liu, L.M., and Hanssens, D.M. (1982). "Identification of Multiple-Input Transfer Function Models," <u>Communications in Statistics - Theory and Methods</u>, 11(3), pp. 297-314.

Mabert, V.A. (1975). "An Introduction to Short Term Forecasting Using the Box-Jenkins Methodology", Publication #2 AII.E. Monograph Series, Production Planning and Control Division, Norcross, Georgia.

Makridakis, S., Wheelright, S. and McGee, V. (1983). Forecasting: Methods and Applications. New York: Wiley.

McCleary, R., and Hay, R. (1980). Applied Time Series Analysis for the Social Sciences. Los Angeles: Sage.

Nelson, C.R. (1973). Applied Time Series Analysis for Managerial Forecasting. San Francisco: Holden Day.

Pack, D.J. (1978). "Forecasting Time Series Affected by Identified Isolated Events," College of Administrative Science, The Ohio State University, Ohio.

Pack, D.J. (1977). "Revealing Time Series Interrelationships," Decision Sciences, 8, pp. 377-402.

Quenouille, M.H. (1957). The Analysis of Multiple Time Series. London: Charles Griffin & Company.
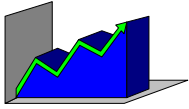
Reilly, D.P. (1980). "Experiences with an Automatic Box-Jenkins Modeling Algorithm," in Time Series Analysis, ed. O.D. Anderson. (Amsterdam: North-Holland), pp. 494-508.

Reilly, D.P. (1987). "Experiences with an Automatic Transfer Function Algorithm," in Computer Science and Statistics Proceedings of the 19th Symposium on the Interface, ed. R.M. Heiberger, (Alexandria, VI: American Statistical Association), pp. 128-135.

Shumway, R.H. (1988). Applied Statistical Time Series Analysis. Englewood Cliffs: Prentice Ha Hall.

Tiao, G.C., and Box, G.E.P. (1981). "Modeling Multiple Time Series with Applications," Journal of the American Statistical Association, Vol. 76, pp. 802-816.

Tsay, R.S. (1986). "Time Series Model Specification in the Presence of Outliers," Journal of the American Statistical Society, Vol. 81, pp. 132-141.

Tsay, R.S., and Tiao, G.C. (1983). "Consistent Estimates of Autoregressive Parameters and Extended Sample Autocorrelation Function for Stationary and Nonstationary ARMA Models," Technical Report #2, Statistics Research Center, Graduate School of Business, University of Chicago, Chicago.

Vandaele, W. (1983). Applied Time Series and Box-Jenkins Models. Orlando: Academic Press.

4
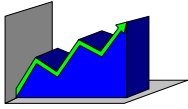
Reviews

Tim C. Ireland and Ramesh Sharda, A Test of Box-Jenkins Forecasting Expert Systems for Microcomputers, Department of Management, Oklahoma State University, College of Business Administration, Stillwater, OK 74078, Phone: (405) 744-8642, (405) 744-8638.

John C. Picket (1991), Autobox 3.0, International Journal of Forecasting, 7:395-398.

John C. Picket (1991), Autobox 3.0, OR/MS Today, pp. 32-35, April 1991.

David P. Reilly (1980) "Experiences with an Automatic Box-Jenkins Modeling Algorithm", in Time Series Analysis, ed. O.D. Anderson. (Amsterdam: North-Holland). Pp. 494-508

David P. Reilly (1987) "Experiences with an Automatic Transfer Function Algorithm", in Computer Science and Statistics Proceedings of the 19th Symposium on the interface, ed. R.M. Heiberger, (Alexandria, Va.: American Statistical Association), pp. 128-135.

J. Keith Ord (1986), Autobox - Software Review, International Journal of Forecasting, 2:512-513.

R.H. Shumway (1986), Autobox (Version 1.02), The American Statistician, 40(4):299-300.

Leonard J. Tashman and Michael L. Leach (1991), Automatic forecast software: A survey and evaluation, International Journal of Forecasting, 7:209-230.

Pamela A. Texter and J. Keith Ord, (1989), Forecasting using automatic identification procedures: A comparative analysis, International Journal of Forecasting, 5:209-215.

Pamela Texter-Geriner and J. Keith Ord, (1991), Automatic forecasting using explanatory variables: A comparative study, International Journal of Forecasting, 7:127-140.