Package “AutoboxRun”

February 28, 2016

Automatic Forecasting Systems©, Inc.

Topics documented:

Set up

If you are having any difficulties please email or call us and we will address your questions immediately. Refer to "Report Errors" section as well and then email us at sales@autobox.com so we can address any issues you might have.

We do recommend suppressing all reports generated which can help run time, by including creating a text file called mute.afs and/or noaux.afs. Try one or the other or both and see the difference in reporting.

The first position in the iObjective allows you start with an assumed model and Autobox will modify the model for outliers, etc if needed. Refer to the Appendix and Objective 1 where we list of the models you can use.

Background

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.

Autobox provides many reports/graphs and early warning system reports along with a verbal description of the model that explain the model in a sentence format. Graphs of autocorrelation, partial-autocorrelation and cross-correlation functions are also available.

When you go to model, you can include causal variables, retain future observations for error analysis, provide future values of the causal variables or tweak the modeling process that Autobox uses.

Autobox will automatically aid the modeling process for weekly, daily, hourly and semi-hourly data. If you have weekly, 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, Autobox will consider 11 monthly dummies and 6 dummy variables for day of the week.

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 at least 78 weeks of data.

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

Note: If you have data that is very different in scale, we 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.

Contact us for any questions: AFS Inc.

P.O. Box 563
Hatboro PA 19040
sales@autobox.com
Phone 215-675-0652

About AFS

AFS has been in business since 1976 and has many business and academic customers. Also check out Chris Chatfield's book where he cites in his 2001 book "Time Series Forecasting" on p. 176 as AUTOBOX BATCH as the package for Box-Jenkins modeling software. AFS was also bestowed the honor (2001) in the text by J. Scott Armstrong "Principles of Forecasting" as the 'BEST DEDICATED FORECASTING PACKAGE' See table 8 (http://www-marketing.wharton.upenn.edu/forecast/paperpdf/Tashman-Hoover%20Tables.pdf).

Overrides

There are two ways to override.   Creating a file on the disk or modifying engine.xxx and saving it as engine.afs(explained later).

For, Type “1” overrides please create a file in the folder and it will be applied For Type “2”, create a file with content. For sparse.afs you might want to put “20”.

FILENAME	TYPE	EXPLANATION
FORCE A MODEL AREA		
USEHW.AFS	1	USE HW AS STARTING MODEL
CLASSIC.AFS	2	ENABLES SIMPLE WEIGHTED AVERAGE
SPARSE.AFS	2	SWITCH TO CONTROL WHAT % ARE ZEROES NEEDED TO CALL A SERIES SPARSE (PCT)
MODELLING OPTIONS AREA		
NOARIMA.AFS	1	IGNORES ARIMA OR DIFFERENCING ON Y OR DENOMINATOR"
NOSTEPDN.AFS	1	NO STEPDOWN - NO REMOVAL OF CAUSALS
NOFIXDAY.AFS	1	DISABLES FIXED DAY OF THE MONTH
NOHOLWEK.AFS	1	SUPPRESS HOLIDAY EFFECTS IF ON WEEKEND
STEPUPDE.AFS	2	MAX # OF OUTLIERS TO BE IDENTIFIED
NOMONFRI.AFS	1	DISABLES MON AND FRI EVENTS AROUND HOLIDAYS
MINCOUNT.AFS	2	MIN # OF OBS TO TEST CONSTANCY OF PARAMETERS
INTERP.AFS	1	ENABLES INTERPOLATION
PUL2STEP.AFS	1	CONVERTS PULSES TO STEPS
RETFIXED.AFS	1	KEEPS ALL FIXED EFFECTS
PUL2SPUL.AFS	1	CONVERT ALL PULSES TO SEASONAL PULSES
NOSEASAR.AFS	1	DISABLES SEASONAL ARIMA
WEEKINMO.AFS	1	SEARCHES FOR WEEK WITHIN MONTH EFFECTS

NOWEEKLY.AFS	1	DISABLES THE SETUP OF WEEKLY DUMMIES
MAXLEAD.AFS	1	ENABLES MAXIMUM LEAD ALLOWED FOR TYPE 3 SERIES
NOINTERM.AFS	1	DISABLES INTERMITTENT DEMAND
NOPULSE.AFS	1	DISABLES PULSES
WEEKPCT.AFS	2	# OF MULTIPLES TO USE FOR WEEKLY DUMMIES(1.5)
NOLEAD.AFS	1	DISABLES LEAD EFFECTS
NOFIXED.AFS	1	NO FIXED EFFECTS
NOMONDUM.AFS	1	DISABLES MONTHLY DUMMIES FOR DAILY DATA
NOHOLDUM.AFS	1	DISABLES U.S. HOLIDAY DUMMIES FOR DAILY DATA
NOINT.AFS	1	DISABLES INTEGERIZATION OF FORECASTS
NOLEVEL.AFS	1	DISABLES LEVEL SHIFTS
NOPARCON.AFS	1	DISABLES CONSTANCY OF PARAMETER TESTING - CHOW TEST
NOVARCON.AFS	1	DISABLES CONSTANCY OF VARIANCE CHECK TSAY TEST
USEDATE.AFS	2	ENTER DATE E.G. '__010105' AND IT WILL BE PREPENDED TO DAILY LEVEL DATA SERIES NAME TO TRIGGER DAILY FIXED EFFECTS, ETC
NOSEASP.AFS	1	DISABLES SEASONAL PULSES
NODATE.AFS	1	MODIFIES SERIES NAME BY ELIMINATING __MODDAYR. STOPS DAILY MODELLING FIXED EFFECTS
NODAY.AFS	1	DISABLES DAILY EFFECTS
MAXLEVEL.AFS	2	MAXIMIM NUMBER OF LEVEL SHIFTS ALLOWED
NODYN.AFS	1	NO DENOMINATOR STRUCTURE FOR CAUSALS
MAKEALLI.AFS	1	FORCES IN ALL CAUSALS
PVALHT.AFS	2	SETS TVALUE FOR OUTLIERS 90

MODEL REUSE AREA

STARTMOD.123	2	STARTING MODEL
USEMOD.AFS	1	ENABLES REUSE OF .MOD FILE
SAVEMOD.AFS	1	ENABLE SAVING OF PASS2MOD.123
MODELNOP.AFS	1	SAVE MODELOUT.ASC without PULSES

REPORTS AREA

NOAUX.AFS	1	DISABLES CREATING AUXILIARY FILES (SAF/ADJ/BOT/BIN
NOCSV.AFS	1	CONTROLS CREATION OF CSV FILES
INTRVENT.AFS	1	SAVES INTRVENT.HTM AS SERIESNAME.INT
FITNDERR.AFS	1	SAVES FIT VALUES AND ERROR VALUES AND BREAKIN/BREAKOUT FILES

FORECASTS AREA

		RESTRICT FORECASTS TO INTEGERS
INTEGER.AFS	1	IVAL=INTF(FORECAST+.5)
FORECONF.AFS	2	SET CONFIDENCE LIMITS FOR FORECAST EXAMPLE 95.0
POSITIVE.AFS	1	FORCES ALL FORECASTS TO BE POSITIVE
NUMBFORE.AFS	1	CONTROLS THE NUMBER OF FORECASTS

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)
Doparcon.txt – Reports the Parameter Constancy test (ie Chow test).
Modified.123 – Series cleansed of outliers
*.tra – Actuals and Forecasts ready for import into tools like Tableau/Spotfire/Qliq

*.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)

*.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

PNG - 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.

Simulated Forecasting – Creating as distribution of forecasts instead of 1 avoiding the “flaw of averages”. Create a file called afs2sip.afs to do so.

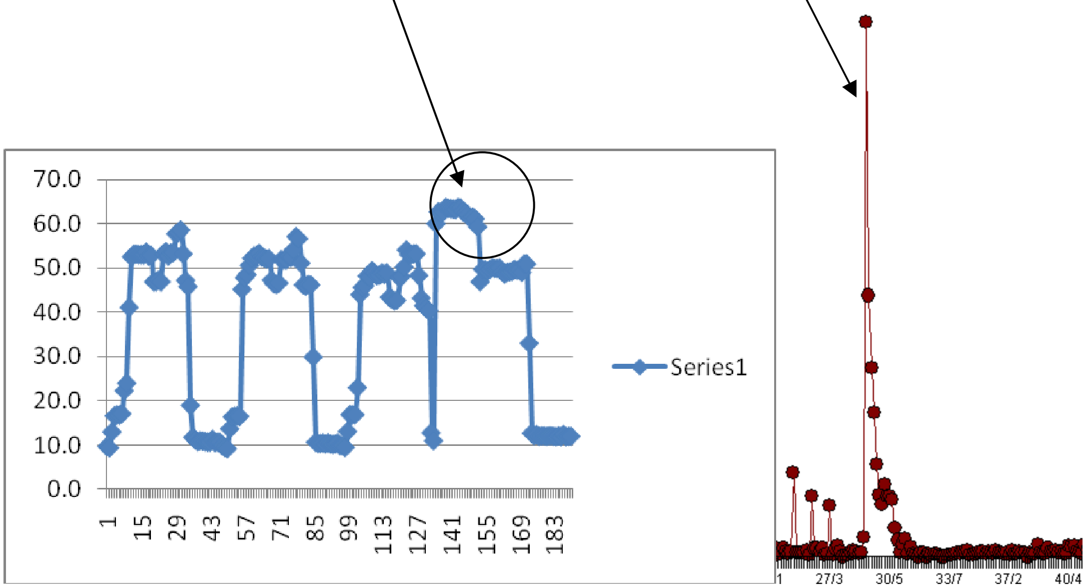
AFS2SIP.afs – Generates Sipmath 2.0 ready XML files(see probabilitymanagement.org for more on their free and very useful simulation tool). Put “1000” in this file and Autobox will generate 1,000 simulations for each of the forecasting periods(see image below). If you put a 1 in the second line, outliers that were identified and adjusted for in the history will be allowed to occur in the forecast and in effect widen the confidence limits as outliers do occur! We recommend using 1,000 as the forecasting confidence limits are in fact generated organically from the data so you want enough trials. If you are forecasting out 36 months then you will have 36,000 forecasts along with the average, standard deviation, min and max. Look for a file called SIPXXX.CSV in the Autobox folder after running Autobox(where xxx=seriesname)


```
<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"
```

Dynamic Modeling

Let's discuss a special way to allow Autobox to model the decay of a promotion like that of an "Ad Stock" approach.

There is one trigger based on the name of the causal variable. If you have a "dynamic" promotion over a period of two weeks (and you have 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'.



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

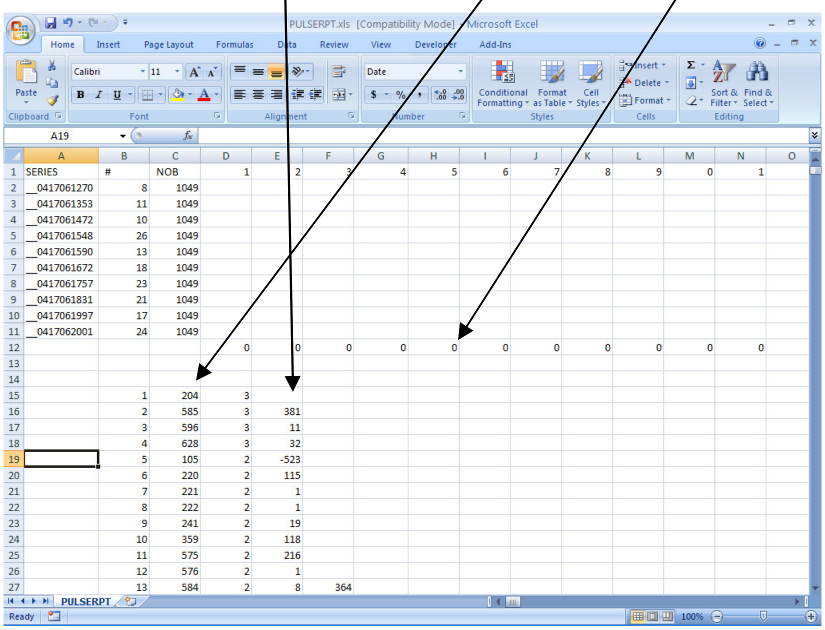
- Forecast Monitoring Report - Quick check to make sure the Autobox forecast is reasonable vs. a simple averaging method of your choosing (ie 4 period equally weighted average) as a baseline for comparison

- Forecasting Fit Exception Report - Compare Autobox fit vs 2 simple methods

Outlier Exception Reports

Pulserpt.csv, Trendrpt.csv, Levelrpt.csv – Log file showing a Table of a pulse(or trend or level) outlier 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 folder. This file is a log file so you can delete it once in a while to start fresh.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	SERIES	#	NOB	1	2	3	4	5	6	7	8	9	0	1	
2	__0417061270	8	1049												
3	__0417061353	11	1049												
4	__0417061472	10	1049												
5	__0417061548	26	1049												
6	__0417061590	13	1049												
7	__0417061672	18	1049												
8	__0417061757	23	1049												
9	__0417061831	21	1049												
10	__0417061997	17	1049												
11	__0417062001	24	1049												
12				0	0	0	0	0	0	0	0	0	0	0	
13															
14															
15		1	204	3											
16		2	585	3	381										
17		3	596	3	11										
18		4	628	3	32										
19		5	105	2	-523										
20		6	220	2	115										
21		7	221	2	1										
22		8	222	2	1										
23		9	241	2	19										
24		10	359	2	118										
25		11	575	2	216										
26		12	576	2	1										
27		13	584	2	8	364									

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 folder NOT the OUTPUT folder.

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 here. However, I went and 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 here. 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

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 folder 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		

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 folder 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, like always you can have 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

Measuring Accuracy

This utility is for evaluating forecasting accuracy from many withheld periods

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 between each withheld amount. You would typically want the second row to be a "1", but if you wanted to get only every other period you could put a "2" for example, but this isn't usual to do. In this example, we used 20 and a 1. We have 144 observations and we have 12 forecast periods so the first observation you can withhold 132 as you can see from the file created "summape.csv" as shown below.

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 HT1 files generated on the disk are the analysis files if you want to review.

Intermittent Demand

Autobox checks if more than 25% of the observations are zero and 6 or more non-zero to zero transitions occur it will then use the following scheme to predict the intermittent demand.

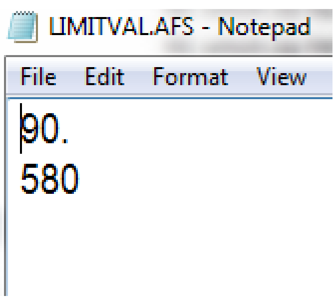
Two new time series will be created: the interval and the rate. The interval series will be the number of periods with zero demand between periods where there is demand (i.e. If there is demand at every time period then the interval would be 1). The rate is the interval divided by the demand.

Autobox will search for level shifts and we find this as a better alternative to Croston's method.

The override file "Sparse.afs" allows you to override the % zeroes to trigger this approach.

Warn Me (What Date We Would Have Need to Build that New Warehouse We Said We Always Needed to Build)

We have added a new feature to Autobox 7.0 that allows a Capacity minded user to define a specified number(ie threshold) and have Autobox tell you at what future date the Forecast's Upper Confidence level will exceed that threshold. When you are trying to determine when your Capacity will be pierced, you can easily do that now. Create a file called limitval.afs and put the statistical confidence level (ie 90.) and the number(580) you want to specify and save the file to the Autobox folder.



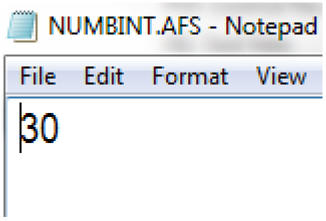
90.
580

After running, a file called Exceed.csv will now exist that tells you the date when the forecast’s Upper Control Limit will exceed that limit that you specified.

	90% FORECAST VERSUS CRITICAL VALUE			
NAME	PERIOD	DATE	FORECAST	CRITICAL VALUE
BJ07	150	1961/ 6	587.0252811	580.0000000000

Tell Me Which Are The Largest Outliers that I Need To Research

We also added a feature to allow the user to specify the number of outliers to be reported to focus only on the LARGEST to research. Sure, Autobox will identify and adjust for your outliers, but what if you only have time to research, identify and correct for a limited amount? You need a way to sift through the models and equations easily and we have done that for you now. If you only want to see the top 5, you get the top 5! Create a file called numbint.afs and specify the number of outliers you want reported and save the file to the Autobox folder.



After running, a file called Numbint.csv will now exist that tells you the date when the forecast's Upper Control Limit will exceed that limit that you specified.

```
* * * Top of File * * *
          30
I~P00050094             0.7519D+03
I~P00006004             0.7357D+03
I~P00037070             0.6035D+03
I~P00016028             0.4779D+03
I~P00015095             0.4361D+03
I~P00015096             0.4349D+03
I~P00016029             0.4317D+03
I~P00016027             0.4278D+03
I~P00016026             0.4085D+03
I~P00016031             0.3952D+03
I~P00016030             0.3850D+03
I~P00015094             0.3811D+03
I~P00018060             0.3759D+03
```

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

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	UNITED KINGDOM
17	AUSTRALIA

If you want to create 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 USA) and make changes to it and save it as ROS18.HLD. You can then create a file called "country.afs" with an "18" in it 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 37250th day since 1/1/1900(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.

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.

<u>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.</u>

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

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

BJ07

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 month 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 summary of the ASC files major sections:

Objectives (all are required)

Data properties (all are required)

Data names (in the order of 1^{st} 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 '0'
OBJECTIVE (3)	No longer used leave as a '0'

DATAPROP Structure:

Name	Description
DATAPROP(1)	Number of series in the problem
DATAPROP(2)	Seasonality. How often the data was sampled. (i.e. Choose 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.) Please note that all series in the model must have the same seasonality
DATAPROP(3)	Major Period -(ie 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)	Minor Period - (ie Beginning month) 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 Minor Period. If you wish to use series whose original Minor Period are different, you must determine the common matrix for the series and use that starting point as the Beginning Period. Here are some examples on the Major/Minor Period –For a monthly data problem let's assume the starting month is February and the year is 2006. The Minor would be a '2' and the Major would be '2006'. For a quarterly data problem let's assume the starting quarter is 3 and the year 1974. The Minor would be a '3' and the Major '1974'. For a daily data problem let's assume the starting week 35 and day 5. The Minor would be '5' and the major '35' . For a weekly problem, let's assume the starting week was the 34th of the year in year 1992. The minor would be '34' and the major '1992'. The Major and Minor are for helpful for reporting purposes only and not for the actual analysis. When we run daily data on test sets, we just put in a '1' and a '1' for each because it takes too much time to figure out the actual week and day so we don't bother with it.
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

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 ‘7’ then daily effects such as “day of the week”, “week of the year”, “day of the month” are analyzed.

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 . Series name must be 8 or less characters 3 = Future values are user specified; contemporaneous, lag and lead effects allowed. Series name must be 8 or less characters 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. 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.

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) beginning 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

You can outright specify what model you want to use. You would need to create a file named "startmod.123". Autobox batch will see that this file exists and then use it. The format shown below is also the same format as the "MOD" files that get created.

Record Layout for: MODEL FORM FILE

NOTE > Make each entry an integer unless it is decidedly real (a model coefficient for example).

Line 1 : contains 5 entries

entry 1: Name of the dependent series columns 1-12 left justified
entry 2: Major Period (example 1991) columns 27-30 right justified
entry 3: Minor Period (example 1) columns 35-38 ..
entry 4: Seasonality (example 12) columns 43-46 ..
entry 5: Number of Series in Model columns 51-54 ..

Line 2 : Enter a zero to indicate that there is not a constant parameter in the model, enter a one otherwise

NOTE > Line 3 is only included when the value of the entry on line 2 is not zero.

Line 3 : Enter the coefficient value for the constant

Line 4 : Contains two values separated by a space or a comma

entry 1: Enter the lambda value (the transformation parameter).

For example , enter a 1 to indicate the original series, enter a 0 to indicate logs, enter a -1 to indicate reciprocals, etc.

entry 2: Enter the offset for the transformation (normally zero)

NOTE > Line 5-9 relate to the ARIMA structure of the model.

Line 5 : contains 3 entries separated by a space or a comma

entry 1: Enter the number of autoregressive factors in the model
entry 2: Enter the number of differencing factors in the model
entry 3: Enter the number of moving average factors in the model

NOTE > Line 6 is only included when entry 2 on Line 5 is greater than 0

Line 6 : Enter order of differencing factor(s) separated by a space or a comma

NOTE >   Skip Line 7,8 and 9 if entry 1 and 3 on Line 5 is zero.

Line 7 : Enter the number of parameters in each of the autoregressive factors specified in entry 1 on Line 5 followed by the number of parameters in each of the moving average factors specified by entry 3 on Line 5.

Line 8 : Enter the backorder powers (or the lag values) for each of the autoregressive parameters , followed by each of the backorder powers for each of the moving average parameters. Enter the lag values factor by factor in ascending order within each factor.

Line 9 : Enter the matching parameter values for each of the backorder powers specified in Line 8.

NOTE >   Line 10-9 relate to the CAUSAL structure of the model.

NOTE >   Repeat the sets of lines 10-16 for each input series. These lines represent the Causal Component in the Transfer Function Model

Line 10 : Enter the name of the input series in columns 1-22 (left justified)

Line 11 : Contains two values separated by a space or a comma

entry 1: Enter the lambda value (the transformation parameter). For example , enter a 1 to indicate the original series, enter a 0 to indicate logs, enter a -1 to indicate reciprocals, etc.

Line 12 : contains 3 entries separated by a space or a comma

entry 1: Enter the number of output lag   factors in the model
entry 2: Enter the number of differencing   factors in the model
entry 3: Enter the number of input lag   factors in the model

NOTE >   Line 13 is only included when entry 2 on Line 12 is greater than 0

Line 13 : Enter order of differencing factor(s) separated by a space or a comma

NOTE >   Skip Line 14,15 and 16 if entry 1 and 3 on Line 12 is zero.

Line 14 : Enter the number of parameters in each of the output lag   factors specified in entry 1 on Line 5 followed by the number of parameters in each of the input lag factors specified by entry 3 on Line 5.

Line 15 : Enter the backorder powers (or the lag values) for each of the output lag parameters , followed by each of the backorder powers for each of the input lag parameters. Enter the lag values factor by factor in ascending order within each factor.

Line 16 : Enter the matching parameter values for each of the backorder powers specified in Line 15.

Examples

Exponential 1- model (Univariate modeling)

```
ZZ                 1998    6   12    1
         0
  1.000000000000000     0.0000000000000000E+000
         0          1          1
1
1
1
.5
```

Time plus Autoregressive 1 model(Causal modeling)

```
YTIME              1998    6   12    2
         1
  10.
  1.000000000000000     0.0000000000000000E+000
         1          0          0
1
1
.2
TIME
  1.000000000000000     0.0000000000000000E+000
         0          0          1
         1
         0
    .2
```

ENGINE.AFS

Listing of ENGINE.XXX which is copied to ENGINE.AFS and edited to override Autobox

```
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

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

This has been disabled.

Line 3

VARIANCE OF THE NOISE SERIES |1.0

If you choose to simulate in line 2 then you will need to specify the amount of variance. The more variance the more randomness in the simulated data. The standard is 1.0.

Line 4

SEED VALUE TO START (0 FOR CLOCK) |0

To start the simulation process, the program needs a starting point to iterate from. If you use the same seed value, then you can replicate your data. If you want persistent random simulated data, then use the "0" to use the time of day which is a pretty good random approximation.

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.