You paid that much and got that little? Do you really know what your forecasting model is doing?

Posted on Feb 15, Posted by Tom Reilly Category Forecasting

We got into an interesting debate on model performance on a classic time series that has been modeled in the famous textbooks and we tried it out on some very expensive statistical software. It's tough to find a journal that is willing to criticize its main advertiser so we will do it for them. The time series is Sales and Advertising and you can even download down below and do this with your own forecasting/modeling tool to see what it does! Feel free to post your results and continue the debate of what is a good model!

What ingredients are needed in a model? We have two modeling violations that seem to be ignored in this example:

1)Skipping Identification and "fitting" based on some AIC criteria. Earlier researchers would restrict themselves to lags of y and lags of X and voila they had their model.

2)Ignoring Modern day remedies, but not all do this. Let's list them out 1)Outliers such as pulses, level shifts, time trends and seasonal pulses. The historical data seems to exhibit an increasing trend or level shift using just your eyes and the graph. 2)Dealing with too many observations as the model parameters have changed over time(ie Chow test) 3)Dealing with non-constant variance. These last two don't occur in our example so don't worry about them right now.

Are you (or by default your software) using dated methods to build its regression model? Is it leaning on the AIC to help build your regression using high order lags? Said more clearly, “Are you relying upon using long lags of X in a regression and ignoring using Stochastic (ie ARIMA) or deterministic empirically identified variables to build your model?” Are you doing this and doing it automatically and potentially missing the point of how to properly model? Worse yet, do your residuals look like they fail the random (ie N.I.I.D) tests with plots against time? The annointed D-W statistic can be flawed if there are omitted dummy variables needed such as level shifts, pulses, time trends, or seasonal pulses. Furthermore, D-W ignores lags 2 and out which ignores the full picture.

See the flow chart on the right hand side on the link in the next sentence. A good model has been tested for nece ssity and sufficiency . It has also been tested for randomness in the errors.

While it is easy and convenient (**and makes for quick run time**) to use long lags on X, it can often be insufficient and presumptory (see Model Specification Bias) and leave an artifact in the residuals suggesting an insufficient model.

Regression modelers already know about **necessity** and **sufficiency** tests, but users of fancy software don't typically know these important details as to how the system got the model and perhaps a dangerous one? Necessity tests question whether the coefficients in your model are statistically significant (ie not needed or "stepdown"). Sufficiency tests question whether the model is missing variables and therefore ruled as insufficient(ie need to add more variables or "stepup").

Is it possible for a model to fail both of these **two** critical tests at the same time?Yes.

Let's look at an example. If we have a model where Y is related to X and previous of values of X up to lag and including lag 4, and lags 2, 3 and 4 are not significant then they should be deleted from the model. If you don't remove lag 2, 3 and 4 then you have failed the necessity test and you model is suboptimal. Sounds like the stepdown step has been bypassed? Yes. The residuals from the "overpopulated" model could(and do!) have pulses and a level shift in the residuals ignored and therefore an insufficient model.

Let's consider the famous dataset of Sales and Advertising from Blattberg and Jeuland in 1981 that has been **enshrined** into textbooks like Makradakis, Hyndman and Wheelwright 3rd edition. See pages 411-413 for this example in the book. The data is 3 years of monthly data.

http://mansci.journal.informs.org/content/27/9/988.full.pdf

Sales -
12,20.5,21,15.5,15.3,23.5,24.5,21.3,23.5,28,24,15.5,17.3,25.3,25,36.5,36.5,29.6,30.5,28,26,21.5,19.7,19,16,20.7,26.5,30.6,32.3,29.5,28.3,31.3,32.2,26.4,23.4,16.4

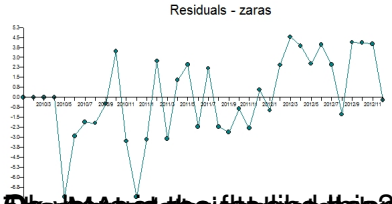
Adv -
15,16,18,27,21,49,21,22,28,36,40,3,21,29,62,65,46,44,33,62,22,12,24,3,5,14,36,40,49,7,52,65,17,5,17,1

The model in the textbook has 3 lags that are not significant and thereby not sufficient. The errors from the model show the need for AR1 when one is not needed of course due to the fact that there is a poor model being used. The errors are not random and exhibit a level shift that is not rectified.

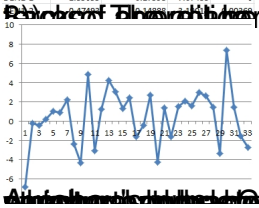
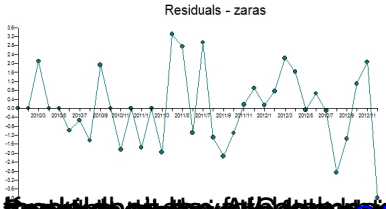
In 2013, the largest company that offers statistical software (and very expensive) is seemingly inadequate. We will withhold the name of this company. The Emperor has no clothes? Nope. She does not. Here is the model in the textbook estimated in Excel (which can be reproduced in Autobox). The results in the textbook are about the same as this output. You can clearly see that lags 3,4,5 are NOT SIGNIFICANT, right? Does that bother you?

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	15.06043374	2.115909576	7.117711416	1.47131E-07
X Variable 1	0.144733983	0.035173523	4.114856067	0.000346304
X Variable 2	0.145184578	0.038373987	3.783411346	0.000820347
X Variable 3	0.042864138	0.038341291	1.117962832	0.273806714
X Variable 4	0.027636725	0.038856874	0.711244156	0.483265631
X Variable 5	-0.026925104	0.036954091	-0.728609555	0.472752738

The residuals are not random and exhibit a level shift in the second half of the data set and two big outliers in the first half not addressed.



Name	Coefficient	Standard Error	T-Values	P-Values
MA1-1	-0.06135	0.17788	-0.34491	0.73294
MA1-2	0.93864	0.20197	4.64739	0.00009
INTERCEPT	0.12187	0.03439	3.54381	0.00152
NUM1-1	-0.29999	0.06389	-4.69527	0.00008
NUM1-2	-0.1977	0.03572	-5.53402	0.00001
DEN1-1	-1.35055	0.17598	-7.67433	0



#	MODEL COMPONENT	LAG (BOP)	COEFF	STANDARD ERROR	P VALUE	T VALUE
1	CONSTANT		13.0	.894	.0000	14.53
INPUT SERIES X1	M_ADV					
2	Omega (input) -Factor # 1	0	.125	.173E-01	.0000	7.22
3		1	-.160	.181E-01	.0000	-8.85
4		2	-.738E-01	.179E-01	.0003	-4.11
INPUT SERIES X2	I-L00026		LEVEL	2012/ 2		
5	Omega (input) -Factor # 2	0	4.17	.666	.0000	6.26
INPUT SERIES X3	I-P00012		PULSE	2010/ 12		
6	Omega (input) -Factor # 3	0	-6.94	1.83	.0008	-3.79
INPUT SERIES X4	I-P00005		PULSE	2010/ 5		
7	Omega (input) -Factor # 4	0	-5.98	1.77	.0024	-3.37
INPUT SERIES X5	I-P00004		PULSE	2010/ 4		
8	Omega (input) -Factor # 5	0	-4.94	1.78	.0099	-2.78
INPUT SERIES X6	I-P00014		PULSE	2011/ 2		
9	Omega (input) -Factor # 6	0	5.09	1.82	.0098	2.79
INPUT SERIES X7	I-P00010		PULSE	2010/ 10		
10	Omega (input) -Factor # 7	0	4.38	1.76	.0193	2.50