Stat 551 Class Project (v1.2)

I want you to do a somewhat realistic group project using SAS FS ©. Unfortunately, we are limited to what can be done through SAS OnDemand using the built-in datasets. So here is the best I have been able to come up with.

There is a built-in SAS dataset called "GULFOIL" that has monthly Gulf of Mexico gas and oil production for a number of areas across 2 regions over a 10 year period. I want you to do forecasting for either "oil" or "gas" for 12 months following the end of the dataset. Whichever variable you choose to model, you may/should use the other series as a covariate series (an "x"). Use only lags of at least 1 for the predictor variable (no fair using tomorrow's oil to predict tomorrow's gas in this context!).

I want you to do the forecasting in a way that honors the hierarchical structure of the problem, using models fit at the area level to inform forecasts at the regional (and overall "project") level. (Set up a hierarchy and reconcile forecasts "bottom up.")

Do this project in groups of at least 2, and ultimately I want a single .pdf report from each group. There are more "areas" than you will have time to "hand-fit" models for (this aspect of the problem is absolutely true to the way this stuff is faced in practice and the reason for the program's generation of potential models). I want you to fit at least 5 area-level models for each group member (and let the "system" provide the other models). My advice is that you choose the areas to model from the set of possible areas so that they represent the largest possible total production. (Look at the individual series and pick the top ones in terms of level near the end of the data set. This mimics what you would do in a business setting where time prevents you from carefully modeling "everything." You would put your time and energy where it should provide the best return in terms of quality of prediction.)

Also, fit 2 region-level models and see how forecasts from those models compare to what you get aggregating up from the "areas."

Use an "SMAPE" model fitting criterion based on an appropriate "holdout." (You should probably think about the implications of using a 4 or 6 or 12 period holdout, given these are monthly data, and given that you have only 10 or 11 years worth of data at your disposal.)

Details for reporting:

1) Have a serious look at the materials posted

http://www.public.iastate.edu/~vardeman/RTGWritingStuff.html
(particularly at the slides of Nell Sedransk and my "Some Injunctions …"). There is also relevant material posted

http://www.public.iastate.edu/~vardeman/IE361/ie361vard.html

under the "Writing Short …" and "some Points of Advice …" And the advice on my Stat 544 page

http://www.public.iastate.edu/~vardeman/stat544/ProjectPointers.pdf

is very relevant. Pay attention to this stuff! For example, DO NOT write in the passive voice!! DO NOT write "they" when you really mean "he or she" (or "she or he"). Make use of proper typography (math font, etc.). All pages, figures, and tables should be numbered. At least in the main body of the report, every figure and table needs a caption and must be referred to in the text (it needs to be there to explain what you are saying in English). All appendices (and subsections of the appendices) must have clear titles. DO NOT leave me guessing about why something is included. Etc., etc. etc.

2) Begin with a title page providing names and e-mail addresses of the group members. Follow that with an abstract summarizing your report in no more than a page. (These don't count against your page limit.)

3) Produce a main body of a technical report of between 5 and 15 pages (exclusive of front matter and appendices, using 11 or 12 point type, 1.5 or 2 line spacing, and 1 inch margins) describing your work and laying out what you've been able to do in terms of both fitting the existing oil or gas values and predicting the "future" ones.

4) Include in an appendix a summary for every model you "hand-build" that provides at least the model form (this can be pasted in from FS), the table of parameter estimates and significance levels, the plot of prediction bounds, and hold-out and whole-sample SMAPEs. (This material doesn't count against your page limit.)

5) E-mail me a .pdf version of your report by 5PM on Friday December 14, 2012.