W.Q. Meeker
1 March 2019

Statistics 451

TIME SERIES ANALYSIS PROJECT GUIDELINES

As an integral part of this course, students are asked to find and analyze their own time series data for two different projects.

- Project 1 Nonseasonal Time Series Analysis
- Project 2 Seasonal Time Series Analysis

For both projects, I require that students work in groups of size two or three (three is the maximum group size).

For both projects, you should find an appropriate data set and proceed to identify, estimate, check, and finally produce forecasts from an appropriate ARIMA (SARIMA for seasonal) time series model. Other analyses are optional. If a leading indicator or an explanatory variable is available for your time series, you may want to fit a transfer function or an intervention model after we study this in class, but this should not be a part of the project that you turn in to me.

For each project, a brief, (two to three pages, not including figures, will usually be adequate) but informative write-up should explain where and how the data were obtained, how the analysis proceeded, including any problems that were encountered (and, how they were solved), results of the analysis and conclusions derived. Suggestions and limitations for forecasting should also be included. The write-up should be done in MS Word and include charts, graphs and/or tables, but *no raw computer output.* Do not include any tables and figures in your project reports unless you refer to them in the text.

1. Choosing Your Data Sets

There are many sources of time series data on the Internet. Government agencies are some of the best sources. But be resourceful. It is better to find a data set is from a process that interests you. Please do not use data taken from time series text books or repositories of text book data such as http://robjhyndman.com/tsdldata/cnelson/cpi.dat http://www.statsci.org/datasets.html. I have seen most of these old "text book" examples. Find something fresh and up to date. Please, do not use temperature or stock-price data for your nonseasonal projects (the models are predictable and not interesting).

2. Project Milestones and Submission

You are asked to submit two project milestones for each project.

- Milestone 1 is a brief proposal giving a paragraph describing the source and background for the data and a plot of the realization versus time.
- Milestone 2 should contain the information in Milestone 1 (to refresh my memory) but also a description of your tentative identification, including plots of the iden output.

The purpose of the milestones is to provide feedback to project teams and to keep them on schedule. You should follow the guidelines given in this document in preparing the milestones. That way, the second milestone can be easily extended into the final project report.

Milestones should be submitted to me as a Word document via email to wqmeeker@iastate.edu (and ***not*** through Canvas) with a copy to your other team member(s), making it easy for me to respond to all with my comments. The name of the Word document should contain the names of all team members in alphabetical order (e.g., Project1_JonesSmithZhang.docx).

Make sure that your final project is self-contained (e.g., it should contain all of the important material from the milestones, perhaps corrected based on feedback) and do not refer back to previous milestones in your final project. Detailed instruction for final project submission are given near the end of Section 6 in this document. Please follow those instructions precisely.

3. Project Grading

Part of your grade will depend on how well you follow the further instructions and guidelines in this document. Specifically, the projects will be evaluated based on the following equally-weighted criteria.

a) Was Milestone 1 turned in on time?
b) Was Milestone 2 turned in on time?
c) Was the problem described adequately in terms of goals of the analysis, source of the data, background of the process that generated the data and so forth. Is there an adequate executive summary?
d) Are there adequate graphics and figures and have they been properly labeled and described in the text.
e) Was the tentative model identification, transformations, and differencing done properly?
f) Was the model fitting done well? Was there adequate discussion of sensible alternative models and an informative model-comparison table?
g) Was the diagnostic checking done properly and completely? Were plot of the ACF of residuals and Q-Q plots of the residuals properly interpreted?

h) Were forecast and forecast intervals compared across models and was there an adequate explanation of differences? Was a hold-out analysis done?
i) Was the writing well organized, clear, and concise?
j) Were the conclusions of the analysis stated clearly?
k) Was the data set and commands submitted on time?
l) Was the project submitted on time?

4. Project 1 Nonseasonal Time Series Analysis

For this project you should find a nonseasonal time series to analyze. You should have at least 40 observations (but more is better, assuming the older data are relevant). For example, you could use daily process data (at least 5 observations per week) for a period of at least 8 weeks (possible examples include electricity consumption (although this might be seasonal, in some applications). Generally, there is no danger of seasonality in yearly data.

5. Project 2 Seasonal Time Series Analysis

For this project the data realization must contain a seasonal (or periodic) component. The realization should contain, if at all possible, at least 50 observations (preferably more) and at least 6 or 7 seasonal periods (but at least 10 or 12 would be better). Seasonally adjusted data must not be used. Note, for example, just because you have monthly data does not mean that the data will be seasonal. For example, for many products and commodities, price data will show little or no seasonality.

For Project 2, use iden( ) to try the four differencing schemes, all combinations of d=0,1 and D=0,1. Include the four graphical outputs in your report. Go to higher differencing only if needed. Beware of over differencing. Report these in milestone 2 and in your final project.

The analysis for Project 2 should be similar to that requested for Project 1. Because your data will be seasonal, however, the choice of an adequate model will probably be more challenging. We will discuss specific guidelines and a strategy for seasonal analysis in class, starting in Module 9.

6. Further Instructions, Guidelines, and Suggestions for the Nonseasonal and Seasonal Projects

a) Briefly describe the process from which your realization came and give the source of your data. If your data is publically available, provide a URL if your data were found on line or specific reference so that someone else could find your data source.
b) Explain the purpose of your time series analysis and modeling.

c) Use appropriate scaling of your data (e.g., use 808.35 million dollars instead of 808350000 dollars).
d) When doing time series analysis for the purpose of forecasting, more data (longer realization) is not always better than less data (shorter realization). Justify your choice for the length of the realization that you use in your project.
e) If you have any "missing observations" explain the cause and explain how you dealt with the problem (for example, it may be appropriate to interpolate to fill in a missing value).
f) Plot your data versus time. Describe important events that occurred over the observation period that probably had an effect on the process.
g) Use the RTSERIES function iden(…) to tentatively identify one or more possible ARMA/ARIMA/SARIMA models for your time series.
h) Use the RTSERIES function esti(…) to estimate at least two different tentative models for your data. Try as many models as are needed until you find one that adequately describes your data, but for your presentation, include the results for no more than 5 different models.
i) Briefly describe the steps that you followed in arriving at your "final" model or models.
j) Briefly describe the behavior of the process that you have been studying and explain what you have learned about this process by studying the data. The written part of this project should be no longer than necessary to convey the important information. Note that conciseness is one of the evaluation criteria.
k) Use iden ( ) to do a range mean analysis to get an indication of whether a transformation is needed or not. Many data sets do not require a transformation.
l) Once you have chosen whether to transform or not, stay with that choice for the complete analysis. Otherwise, modeling becomes too complicated. You only need to include one or two range mean plots in your final report (original data and transformations, if one was used or seriously contemplated).

   If there was some question about whether a transformation is needed or not or what transformation to use, try alternatives after the initial analysis is complete. Plot of residuals versus predicted values and a comparison of predictions intervals are particularly useful in deciding whether to transform or not and what level of transformation to use.

m) Give your data files meaningful names (e.g., MilkPrice.tsd) instead of short abbreviations (e.g. m.tsd)
n) Prepare a table to compare the models that you have fit. Compare at least 3 plausible models. If you fit more than 5 models, present the results of the 5 best models. You should not report the results of models that perform poorly (I display poorly performing models in class for pedagogical reasons, but they have no place in a report). Do not report too many digits in your numerical results in the table. Typically 3 or 4 is enough. Extra

digits only clutter. I find concise summaries of comparative analyses to be most useful and informative.

o) Include the esti( ) graphical output for the most interesting contrast between two models. Give esti( ) graphical output for only these most interesting two models. It is most interesting to compare two models that are plausible, but that give different results.

p) Compare forecasts for the best two models in your set. Briefly comment on any differences between the forecasts for these two models. When comparing two plots, the axes should be the same. Use an argument like y.range=c(100, 1100) to the esti() function to control the axes, making it easier to do comparisons. If you do a hold-out analysis, be sure to make your comparisons in the context of the predictions intervals.

q) Reports should not include any raw computer output. Include only relevant figures and computer output. Note that such output is relevant *only if* you refer to it in your write-up.

r) Divide your report into several logical sections and properly title and number each section.

s) Your reports should be ***concise*** but complete. Your project will be graded on accuracy, completeness, relevance of your graphical displays, tabular comparisons, and your statements, and on the report's organization, conciseness and neatness.

t) Whenever possible, use graphs plus a few words to make your points. Number your figures with informative captions and refer to them in your write up. In Word you can use right-click, Insert Caption and in the caption give a brief description of what is in the figure.

u) Figures and tables should be integrated into the test and ***not placed at the end*** of the report. You can control the size of the figures, but do not make them so small that the axis labels are difficult to read. In tables, avoid scientific notation like 1.631e-03 and use 0.00163 instead. For p-values < 0.001, use < 0.001 as the exact magnitude of smaller numbers has little meaning.

v) Note that Word has capabilities to do simple formulas with Greek letters, subscripts, etc. Learn how to use these capabilities and use them in your project (Insert -> Symbol).

w) The first page of your project report should be a title page giving the title (choose a meaningful title that related to the data, not something like "Project 2") of the report and names of the team members.

x) The second page of your project report should be an Executive Summary that is no more than one page long, giving the title, data set name, data source, final model, and a brief description of your conclusions. Conclusions should contain the stated purpose of your analysis. The conclusions (possibly expanded a little) should also be included at the end of the main report.

y) You should submit your final project electronically (i.e., via email to wqmeeker@iastate.edu) in the Word format. The name of the Word document should contain the names of all team members in alphabetical order (e.g.,

Project1_JonesSmithZhang.docx). You also need to submit your .tsd object. To do this, from R, give the command

> **dump("xxx.tsd", "xxx.tsd.R")**

where xxx.tsd is the name of your data object in R. Then along with the Word document, also email the resulting xxx.tsd.R file and an xxx.txt file that contains the iden() and esti() or other R commands that you used in your project, in a logical order (so that I can reproduce your results if need be). Make sure that the commands that you send correspond to the analyses in your report and make sure that the name of the tsd object in the tsd dump agrees with what you use in your commands. This commands file should ***not*** contain any raw computer output or extraneous clutter (do not just send the R history without careful editing).

When you submit your project via email, everything you submit should be attached to *one single email* and all other team members should be added to the CC line. If I ask you to resend something, resend ***everything*** again and again ***copy your team members.***

z) If you use JMP for your analyses, submit the JMP Data Table with analysis Scripts for all of the relevant analyses embedded into the Data Table. If you use JMP, you should still submit a tsd object containing your data (as described in the previous guideline).

7. Additional Instructions, Guidelines, and Suggestions for the Nonseasonal Project

a) Look carefully at your ACF function to see if there is evidence of seasonality. For Project 1, if there is evidence of nonstationarity, try using (d=1) to do this check. If there is seasonality in your data, you will need to find a new series to analyze.

8. Additional Instructions, Guidelines, and Suggestions for the Seasonal Project

a) For your seasonal project, I suggest data consisting of 12 months per year or 4 quarters per year or 5 (or 7) observations per week. I suggest avoiding long seasonal periods (like 52 weeks per year). If you do have a long seasonal period, you will need to use special arguments in the calls to iden and esti to get more lags in the ACF and PACF functions.

b) For your seasonal project (Project 2) choose one or two differencing schemes (usually (d=0, D=1) and/or (d=1, D=1)) and tentatively identify models. Fit these models using esti( ).

c) When writing your model in shorthand fashion use something like SARIMA$(0,1,1)(0,1,1)_{12}$, including the subscript indicating the seasonal period.