Author's response to reviews

Title: GAMwithAR applied to temperature-mortality study

Authors:

Lei Yang (yanglei_cq@126.com)
Guoyou Qin (gyqin@fudan.edu.cn)
Naiqing Zhao (nqzhao@fudan.edu.cn)
Chunfang Wang (cfwang@scdc.sh.cn)
Guixiang Song (gxsong@scdc.sh.cn)

Version: 5 Date: 2 August 2012

Author's response to reviews: see over
Reply to Michael Hayn
Thank you very much for your review suggestions.
Following content in bold are my corresponding modifications and replies.

chapter "Background":
p. 4, l. 17: Please replace "humility" by "humidity".
The word is replaced by humidity

p. 4, l. 24 (previously p. 5, l. 7): "In other words ...": Also the usaged of lag effects were mentioned when describing the models above. Here, the data can not be rearranged randomly.
REPLY:
>>the wrong sentence is deleted.
NEW REVISION:
The sentence was replaced by "In other words, in these models, the information of how dependent variable relies on its past values is not used." This sentence is still wrong, since the usage of lag effects were mentioned when describing the models above.

Lag effects explain how the dependent variable (mortality) is dependent on past values of other explanatory variable, like time, temperature and humidity. But how the dependent variable depends on past values of itself is not contained in previous modeling framework. Our work is intended to explain this.

chapter "Methods":
p. 6, chapter Methods (previously p. 7, chapter GAMwithAR):
The formulas still require a more detailed description. Thus, in the first formula of the chapter "Methods", "y" should be declared to be the described observable, "X_{ti}" the matrix containing the explaining variables (or covariates) that enter the model linearly by means of the model parameters "beta", and "g" the link function that was introduced at p. 3, l. 24. Please refer to this explanation of the link function.
(Of course, this information can be found in the paper of Benjamin et al., but a published paper should contain an explanation of all the used terms, symbols and abbreviations.) Similarly, for the first formula in chapter "Algorithm", the single factors "f(...)" should be introduced as conditional likelihoods (also for avoiding confusions with the functions "f" mentioned in the chapter "Natural cubic spline") depending on parameters "\beta". In order facilitate the reader the search for these parameters in this special model, You should shift the specification of "\beta" from the chapter "Newton's method" to the chapter "Maximum Partial Likelihood Estimator (MPLE)"
or point to the specification in the chapter "Newton's method".
Thank you very much for your detailed comments. I made the following modifications on your suggestions.
In chapter "GARMA"
Illustrations of the model is added before the formulas. These are mainly adapted from the paper [33].
In Chapter "MLE"
The conditional likelihoods are introduced before the formulas.
The specification of “beta” and "g" is shifted to Chapter “MPLE”.
In Chapter “natural cubic spline”
To avoid confusion, “f()” which is used to refer to natural cubic spline function is replaced by “ns()”, this change also corresponds to the terms used later.

p. 9, l. 12:
please replace "L" by dots. Also at p. 11, l. 3 and 5
Sorry, I reviewed p. 9 l.12, but didn’t quite understand what you mean.

chapter "Evaluation of GAMwithAR by simulation":
p. 12, l. 10 (previously p. 11, l. 23):
In order to explain, what the time points stand for, I suggest to supplement "time points 1828-3288 (year 2005-2008) to eliminate the impact of starting value"by"time points 1828-3288 (year 2005-2008) to eliminate the impact of the startingvalue, where the time points stand for the days since Jan 1st 2000."
On your suggestion, I add this sentence, it is clearer now.

p. 12, l. 13 (previously p. 12, l. 1):
I suggest to supplement"the ACF and PACF[27] of Pearson residual [16]," by"the autocorrelation function (ACF) and partial autocorrelation function (PACF) [27] of the Pearson residuals [16],"
The current sentence is :” For two models on Sample 1, the ACF (AutoCorrelation Function) and PACF(Partial AutoCorrelation Function)# of Pearson residual #, plotted against lags, was given as indicators of residual autocorrelation.”

p. 13, l. 17: "are plotted in figure": Please supplement the figure number, and replace "figure" by "Figure" in order to stay conform to the other table and figure references.
The “figure” is changed to be “Figure 5”

p. 13, l. 23: Please replace "humility" by "humidity".
The correction is made

p. 14, l. 3: "presented in ...": Please supplement the table and figure numbers, and replace "table" by "Table" and "figure" by "Figure" in order to stay conform to the other table and figure references.
These numbers are added.

chapter "Results":
p. 14, l. 10 (previously p. 13, l. 22): Please replace “\phi=2.675 \neq 1” by “\phi=2.675 > 1”.
This change is made.

p. 16, l. 10 (previously p. 15, l. 18): What are “Pearson coefficients”?
If You have in mind the "Pearson residuals", please use this name instead of "Pearson coefficient". Otherwise, please give a definition of (or reference for) the Pearson coefficients.
Pearson coefficients should be Pearson correlation coefficients. To explain why we calculate Pearson correlation coefficients here, I made the following
revision:
In p 13 Chapter simulation study 2
 “To compensate for the absence of true parameters and bias of estimated parameters, we compare two models visually in Figure 5, in which two splines based on average estimated parameters and the true effect are plotted. What’s more, the Pearson correlation coefficient of fitted linear predictors from the two models and true effect are also calculated, a higher coefficient is desirable because it means that the fitted value is more dependent on the true value.”
And in p. 16, l. 10, “Pearson coefficients” are changed into “Pearson correlation coefficients”

chapter "Conclusions":
p. 21, l. 12 (previously p. 22, l. 5): "GAMwithAR, for fitting time series data with covariates"
As explained above (p. 2, l. 5), the term "covariates" does not explain the specifique nature of the current model. Proposal: "GAMwithAR, for fitting time series data with autoregressive effects".

REPLY:
>>covariates refer to other variables, like temperature, not AR.
>>We use this phrase to stress that our model is different from
>>models of univariate AR model which current value y only
>>relies one its past values.

NEW REVISION:
Thank You for the response. Nevertheless, the special purpose of the GAMwithAR is to respect also the autoregressive effects. For fitting time series with covariates, the more simple GAM would be sufficient. The conclusion of the article should summarize and emphasize the real aim and gain of this work. For this reason, I continue suggesting the modification in this sentence.
On your suggestion, this sentence has been changed into “This article proposes GAMwithAR, for fitting time series data with explanatory variables and autoregressive terms.” This modification stresses our model are both a time series with covariates and a GAM with autoregressive terms.

There are still citations that need to be completed. examples:
p. 27, reference 7: volume and pages
This literature is deleted from reference
p. 24, reference 10: "Peng"; The full citations of books also include the publisher
The publisher is added
p. 26, reference 35: name, year, university

They are added

New reference 34: we can’t find the author for this literature, so we left it as blank, and only cites “SAS Institute inc.”
Reply to Xiangming Fang

Thank you very much for your review suggestions.

Following content in bold are my corresponding modifications and replies.

1) Since the methodology does not include the estimation of the degrees of freedom of the natural cubic splines, it would be helpful if the authors could investigate the effect of misspecification of the degrees of freedom through a simulation study. If that is too much work, the authors should at least provide some general ways or guidelines to select appropriate degrees of freedom in practice.

This is a very good comment. In Chapter “Application to the real data”, the way to choose the degrees of freedom is illustrated. The degree of time spline is chosen in a somewhat arbitrary way, while the degrees of other splines and lagged days are determined by minimizing AIC. There are several other methods to determine the degrees of freedom. There is a new paragraph in Chapter “discussion”discusses this topic.

2) For the models defined by equations (1) and (2) on page 7 and the first equation on page 8, there is a grand intercept \( \alpha \). Should \( \alpha \) be included in the autoregressive term as well?

On your suggestion, also to be consistent of the formulas what came before and what follows, we drop intercept in equations in page 7 rather than add intercept \( \alpha \) in the following context.

3) There is a question mark after the estimated value of the dispersion parameter on page 14. Should it be “>”?

The “?” should be “>”, and it is changed to be so.

4) Proofreading by a native English speaker may be needed before publication.

Finally, I have a discretionary revision suggestion. The authors discussed the differences between GAMMs and the proposed GAMwithAR model. They do have different forms in model specification, but in practice they are essentially dealing with the same issue, i.e., GAMs with correlated data. The question is whether you want to model the autocorrelation on the regression errors (GAMwithAR) or on the random effect (GAMM). In the simulation study, the authors compared the results from GAMwithAR to the results from GAM. It may be more fair to compare GAMwithAR with GAMM. It would be more interesting if the authors can add the results from GAMM fitting in the first simulation study.

Thank you for your suggestion. As you said, GEE or GLMM/GAMM are all dealing with correlated data, so they seem to be very similar to GAMwithAR. And it may be necessary to compare GAMwithAR to GAMM in simulation. However, mixed type model deals with grouped data; observations in different cluster need to be independent while observations within a certain group can have a particular correlation structure. You can refer to [1] to see such difference, “Mixed-effects models are primarily used to describe relationships betweena response variable and some covariates in data that are grouped accordingto one or more classification factors. Examples of such grouped data includelongitudinal data, repeated measures data, multilevel data, and block designs. By associating common random effects to observations sharing the samelevel of a classification factor, mixed-effects models flexibly represent
the covariance structure induced by the grouping of the data.” (P3, 2nd paragraph)

For GLMM and GAMM, refer to [2]: “Generalized linear mixed models follow from linear mixed models, as GLMs followed from linear models.” (P303, Section 6.4) “A GAMM is just a GLMM in which part of the linear predictor is specified in terms of smooth functions of covariates” (P309, L1). So GAMM is also used to describe grouped data.

More specifically, we simulate two sets of data from GLMM and GAM with AR to show their difference.

For a GLMM:

\[ y_i \mid b_i \sim \text{Poisson} \]
\[ \ln(E(y_i \mid b_i)) = 10 + b_i \]
\[ b_i \sim N(0, G) \]
\[ G = 4 \times \begin{pmatrix} 1 & 0.5 & 0.25 \\ 0.5 & 1 & 0.5 \\ 0.25 & 0.5 & 1 \end{pmatrix} \]

\[ i = 1, 2, 3 \]
\[ y_i = (y_{i1}, y_{i2}, y_{i3}) \]
\[ b_i = (b_{i1}, b_{i2}, b_{i3}) \]

For a GAM with AR:

\[ y_t \sim \text{Poisson} \]
\[ \ln(E(y_t)) = 10 + 0.5 \times (\ln(y_{t-1}) - 10) \]

\[ t = 1, \ldots 9 \]

Data set 1 from GLMM:
(8522, 9961, 11881), (84345, 75832, 112687), (2500, 18046, 13955)

Data set 2 from GAM with AR:
22189, 21849, 21860, 21894, 21890, 22078, 22247, 22258, 22342

For data set 1, \( y \) are correlated for every 3 data points, so they are a cluster, but \( y \) are independent between clusters, and the disparities between every cluster are dependent on variance of \( b \), here is 4. For data set 2, \( y \) are serially correlated.

Due to different data generation mechanism, they cannot be compared in the
simulation when evaluating the performance of GAMwithAR. And to avoid confusion, we deleted the former chapter Appendix.

Reply to editors and both reviewers
Thank you very much for your review suggestions. The article is revised with the help of some experienced writers and please let me know if the English writing quality need to be further improved.

References: