

Response to Reviewer:

I am grateful for the time and effort the reviewer has devoted to the paper. The comments made by her/him have undoubtedly resulted in an improved paper. Below I detail how I addressed each of the comments.

Comments

1. *Table 1. What defines ‘User-specified’? For example, if I select the ‘Advanced Autometrics settings’ in Autometrics I can choose from a set of diagnostic tests, and under ‘Tie-breaker’ I can specify a goodness of fit measure. I am not familiar with Grocer to know if I can make similar choices. Would these specifications not count as ‘User-specified’ in your setting?*

No, they would not. At suitable places I have added “in a pre-programmed way” to make it clearer what is meant by “user-specified”.

2. *Equation (2): do you need to provide any assumptions on ϵ_t ?*

No, since the assumptions required on the error ϵ_t will be determined by the user-specified estimator/model.

3. *p.3, step 1: The diagnostic tests are only feasible when $k < n$. The explanation of ISAT methods doesn’t explain how the diagnostic tests are handled.*

For some user-specified diagnostic tests, $k < n$ is not required. In the `isat()` function, the default is that no diagnostic tests are undertaken (these have to be turned on, or user-specified, by the user). I have added a sentence to make this clearer.

4. *The paper states that the `getsFun` function accepts up to 25 arguments. The four key arguments are well explained in the paper, but it would be useful to list all 25 arguments for the user to modify, perhaps in an appendix, so the extent of user-specification is clear.*

I have added a sentence that refers the reader to [Pretis et al. \(2018\)](#) – which discusses the most important other arguments in the context of the `getsm()` function, and to the help-pages of the `getsFun()` function (available by typing `?getsFun`), for the details of all the arguments.

5. *p.4, path output. Would the regressors be listed with names if the dataset regressors have names assigned, or would the regressors be numbered by order as given in the output in the paper?*

No, the regressors would not be listed with names if the regressors were named in the dataset in question. This is to minimise the print. However, if `paths[[1]]` contains the path 1 deletions, then a name-representation is readily obtained with the command `colnames(x)[paths[[1]]]` or similar. I have added a sentence that states this explicitly.

6. *Could all the estimators that could be used be listed, perhaps in an appendix? The same applies to the diagnostic tests and goodness of fit measures. If too many, perhaps list the most common ones that users are likely to implement.*

No, since the number of possible functions is limited by the user's imagination only. Table 1 provides a summary of the pre-programmed estimators/models available. Regarding the pre-programmed diagnostic tests available in the R package gets, I have added a footnote that indicates how they can be invoked from `getsFun()`. For the pre-programmed goodness-of-fit measures in the R package gets, I have added a small sentence that refers the reader to the help pages of the function `infocrit()`.

7. *p.6-8. R code is reported in three sections, first to specify an alternative estimator, second to specify alternative diagnostics, and third to specify an alternative goodness of fit measure. To aid the reader, the full code could be reported once, highlighting the modifications needed for each of the user-specified changes.*

In an early version of the paper I did in fact choose this approach. A drawback with this approach, however, is that it entails more implicitism. Given that the paper is intended to be an introduction to how user-specified functions have to be structured, I believe the chosen explicitism facilitates understanding.

8. *p.7, setting diagnostic test significance levels. It may be helpful here to add a note of caution on multiple testing to guide the user as to the appropriate choice of significance levels. Given that both the number and size of diagnostic tests can be set by the user, some guidance would be helpful.*

Good point. I have added a footnote that provides guidance.

9. *p.8, turbo. Can you give some indication of comparative computing time with turbo on and off for various specifications? This could guide the user as to when to implement turbo or not. Similarly, by how much does setting parallel on help with computational speed? Can you give some examples to guide the user?*

The speed gain of turning the turbo on depends, amongst other, on the estimator in question, the number of variables and the number of observations. I have added some general guidance on when it will result in faster search. Similarly for parallel computing.

10. *p.10, user specified significance level. Can you show how to set the significance level in the code?*

This is already done in the sentence "...the significance level can be lowered to, say, 1% by setting `t.pval = 0.01`". I have made minor changes to the text prior to this sentence to make this clearer .

11. *p.10. For the uninitiated, could you explain what the S3 refers to in the section header?*

S3 methods are a fundamental part of R, so I have added a pointer to where the uninitiated reader can learn more about S3 methods.

12. *Could you explain how the intercept is handled, including whether to include as fixed or selected over?*

The `lm()` function includes the intercept by default, and I have added a small sentence that explicitly states this. Due to the way the code-example has been put together, whose primary motivation is pedagogical, the intercept is excluded from deletion (i.e. it is "fixed"). I have also added a sentence that states this explicitly.

13. *Regression with ARMA error. Why are only 11 regressors searched over when ISAT is designed for this specific problem?*

The reasons are pedagogical. First, using the `arma()` function in combination with `getsFun()` rather than `isat()` makes the coding simpler. Second, many readers are likely to not be familiar with ISAT methods. Accordingly, the example serves as a gentle introduction to the ISAT example that comes after.

14. *p.12. For the uninitiated, could you explain what ‘ols with method=3’ means?*

It means OLS with the ordinary coefficient-covariance, see the help pages of the `ols()` function. I have added a small sentence that states this explicitly.

Minor Corrections

All the minor corrections have been implemented as suggested.

References

Pretis, F., J. Reade, and G. Sucarrat (2018). Automated General-to-Specific (GETS) Regression Modeling and Indicator Saturation for Outliers and Structural Breaks. *Journal of Statistical Software* 86, 1–44.