

View Letter

[Close](#)

Date: May 08, 2018
To: "tao huang" t.huang@surrey.ac.uk,sphere082175@hotmail.com
cc: ;d.soopramanien@lancaster.ac.uk;r.fildes@lancaster.ac.uk
From: Ruud Teunter eesserver@eesmail.elsevier.com
Reply To: Ruud Teunter r.h.teunter@rug.nl
Subject: Your Submission EJOR-D-18-00185

Ref.: Ms. No. EJOR-D-18-00185
Forecasting Retailer Product Sales in The Presence of Structural Breaks
European Journal of Operational Research

Dear Dr. tao huang,

the reviewers have different views on the potential contribution. For this reason, I waited some extra weeks for the comments of a third reviewer. However, despite reminders, the third reviewer did not deliver and so I base my decision on the two reviews to avoid further delay.

I do think that there is a good contribution here, also because the effect of promotions on retail sales is understudied, certainly in the OR literature. Furthermore, both reviewers offer many suggestions for better positioning.

It has been decided that your paper should be reconsidered for publication in the EUROPEAN JOURNAL OF OPERATIONAL RESEARCH after a major revision taking into account the enclosed referees' comments. (You may wish to argue that some comments are invalid). Please also ensure that you have cited recent and relevant publications in EJOR and other OR journals.

When resubmitting the paper, please provide your answers to the referees, separately for each referee.

To submit a revision, go to <https://ees.elsevier.com/ejor/>
Your username is: *****

If you need to retrieve password details, please go to:
http://ees.elsevier.com/EJOR/automail_query.asp

NOTE: Upon submitting your revised manuscript, please upload the source files for your article. For additional details regarding acceptable file formats, please refer to the Guide for Authors at:
<http://www.elsevier.com/journals/european-journal-of-operational-research/0377-2217/guide-for-authors>

When submitting your revised paper, we ask that you include the following items:

Response to Reviewers (mandatory)

This should be a separate file labelled "Response to Reviewers" that carefully addresses, point-by-point, the issues raised in the comments appended below. You should also include a suitable rebuttal to any specific request for change that you have not made. Mention the page, paragraph, and line number of any revisions that are made, and highlight these parts of the text in the revised manuscript.

Manuscript and Figure Source Files (mandatory)

We cannot accommodate PDF manuscript files for production purposes. We also ask that when submitting your revision you follow the journal formatting guidelines. Figures and tables may be embedded within the source file for the submission as long as they are of sufficient visual quality. For any figure that cannot be embedded within the source file (such as *.PSD Photoshop files), the original figure needs to be uploaded separately. Refer to the Guide for Authors for additional information.

Highlights (mandatory)

Highlights consist of a short collection of bullet points that convey the core findings of the article and should be submitted in a separate file in the online submission system. Please use 'Highlights' in the file name and include 3 to 5 bullet points (maximum 85 characters, including spaces, per bullet point). See the following website for more information
<http://www.elsevier.com/highlights>

Graphical Abstract (optional)

Graphical Abstracts should summarize the contents of the article in a concise, pictorial form designed to capture the attention of a wide readership online. Refer to the following website for more information:
<http://www.elsevier.com/graphicalabstracts>

You will find your paper in 'Submissions Needing Revision'.

I am most grateful for your interest and hope that you will not be deterred from resubmitting this paper in the very near future.

You are advised to revise and resubmit your manuscript by the Nov 04, 2018. After this time the review process of this submission will be terminated, by withdrawing your paper from the Elsevier Editorial System, and the revised paper will have to be submitted as a new submission.

Please note that this journal offers a new, free service called AudioSlides: brief, webcast-style presentations that are shown next to published articles on ScienceDirect (see also <http://www.elsevier.com/audioslides>). If your paper is accepted for publication, you will automatically receive an invitation to create an AudioSlides presentation.

Data in Brief (optional)

We invite you to convert your supplementary data (or a part of it) into a Data in Brief article. Data in Brief articles are descriptions of the data and associated metadata which are normally buried in supplementary material. They are actively reviewed, curated, formatted, indexed, given a DOI and freely available to all upon publication. For more information and a submission template please refer to the Journal's Guide for Authors. Questions? Please contact the Data in Brief publisher, Paige Shaklee at dib@elsevier.com.

European Journal of Operational Research features the Interactive Plot Viewer, see:
<http://www.elsevier.com/interactiveplots>. Interactive Plots provide easy access to the data behind plots. To include one with your article, please prepare a .csv file with your plot data and test it online at <http://authortools.elsevier.com/interactiveplots/verification> before submission as supplementary material.

Include interactive data visualizations in your publication and let your readers interact and engage more closely with your research. Follow the instructions here: <https://www.elsevier.com/authors/author-services/data-visualization> to find out about available data visualization options and how to include them with your article.

Sincerely yours,

Ruud Teunter
Editor

Reviewer comments:

Reviewer #1: This is a nice paper that investigates the influence of structural change (see one point I make below) on retail data in the IRI data set. It proposes to augment established Autoregressive Distributive Lag (ADL) models, by either fitting multiple models to shorter and shorter terminal segments of the series, then averaging the forecast (so data points after the conjectured structural break are weighted more heavily), or by attempting to estimate a bias from the terminal in-sample fits and correcting for this estimated bias. The forecast accuracy is improved in either way. I have few important points to make and mostly recommend toning down some overly enthusiastic claims.

Medium points:

- Highlights: "Retailer product sales from a wide range of product categories" is not very useful.
- "Structural breaks" suggests a sudden and abrupt change in a parameter, like a step change in the overall level of the time series. I find this choice of words somewhat unhappy. On the one hand, in a retail environment, I would typically not expect a sudden abrupt change (except in exceptional cases, like a new store opening next door), but rather a gradual one, which would also be more consistent with drivers of such change that the authors discuss, e.g., shifts in lifestyle. On the other hand, there is a vast literature on detecting structural changes in time series (e.g., look at the documentation of the strucchange R package on CRAN), but the authors nowhere apply such a test (which would not be overly useful in my opinion, per above). Thus, maybe "structural change" instead of "structural break" would be a more useful word for the concept.
- The EWC is very similar to estimating a single model with weighted observations (or residuals), with the weights decreasing as we go further into the past. I wonder whether a direct approach like this would be feasible. It would have the advantage of returning a single parameter vector that could be examined and analyzed, whereas in EWC, one has to look at multiple such vectors, one from each of the models we average.
- None of the accuracy measures used is minimized in expectation by an unbiased forecast (see the introduction of Kolassa, 2016, IJF). I am somewhat concerned that the increase in accuracy may have come at

the cost of bias. I would recommend using either a measure that is minimized in expectation by an unbiased forecast, e.g., a scaled RMSE, or assessing bias in some way.

- I see no discussion of how the estimation was actually carried out. Was some third-party software used, e.g., an R package? Please cite software and packages used (with version numbers). Software authors are entitled to recognition.

- p. 11: what is "four-week seasonality"? Does this refer to cutting the year into 13 four-week periods ($13 \times 4 = 52$), then expending 12 dummies? If so, this corresponds to fitting a periodic *step* function, which stays constant for four weeks and then may change sharply when we enter the next four-week bin. This is discretizing a continuous variable (namely, time). Don't do this. The disadvantages of discretization have been well documented over the years - apart from the almost certainly ecologically invalid step fit I discuss, it also expends far too many degrees of freedom. Instead, use a number of periodic spline transforms of time. Three or four parameters expended here would be much better invested than twelve in discretization. Further reading here:

<https://stats.stackexchange.com/questions/230750/when-should-we-discretize-bin-continuous-independent-variables-features-and-when>

<https://stats.stackexchange.com/questions/41227/justification-for-low-high-or-tertiary-splits-in-anova> (see the links in Glen_b's answer)

- p. 19: To be honest, I do not find the introduction and discussion of the ADL-EWC-IC model convincing, since it was created after analyzing the performance of the separate models on subsamples. It is always easy to build a new model ex post that appears to perform well, but this is little better than data snooping. Can the ADL-EWC-IC model be compared to the other models on previously unseen data? If not, please label this discussion explicitly as exploratory. Similarly, please revisit the discussion in the first paragraph on p. 25.

- p. 23: In discussing a statistically significant positive coefficient for "Randomness and growth" in ADL-intra-EWC, the authors write that "This suggests that our proposed models tend to be more advantageous for the SKUs which are difficult to forecast and exhibit a trend in sales". I do not understand this. The dependent value modeled is MASE, so a positive coefficient of 0.4 should mean that a unit increase in "Randomness and growth" should be associated with a 0.4 unit (up to multiplication by 100, per the footnote to Table 7) *increase* in MASE. That is, ADL-intra-EWC should perform *worse* than the benchmark for high-"Randomness and growth" series, not *better*. Please clarify. Same for the rest of the discussions on the same page, and the third paragraph on p. 25.

- p. 23, "All the results here indicate that we may pre-test these features for each SKU and then determine the optimal sales forecasting method specifically for that SKU." This is a very exploratory finding, not guided by prior hypotheses. Unless this can be verified on previously unseen data, please emphasize the tentative nature of this recommendation.

- p. 24, "The improved forecasting accuracy for product sales substantially contributes to retailers' profit". This is a very strong claim, and not backed by anything the authors did in their paper. Whether forecast accuracy improvements actually translate into better profits is not this clear-cut, since forecasts must still be translated into operational plans, which are constrained by logistics. In addition, supply chain operations rely far more on quantile forecasts than on point forecasts, since the total order includes safety amounts, and whether these quantile forecasts can be improved by the authors' proposals is not obvious. Please tone this claim down.

Minor points:

- p. 6, 1st equation: there is no intercept in the formula. If u_t is assumed to have mean 0, this implies that a price of $x=0$ is associated with sales of $y=0$. I assume the authors meant to include an intercept parameter.

- Table 3: why do different models serve as benchmarks?

- Figure 3: please provide more information in the figure caption instead of in the text or even in footnotes - the reader should not need to hunt through the text to understand the figure. Are the diamonds joined by lines group means? Are box widths *proportional* to numbers of SKUs in each category? "Are determined by" can be a log transform, a square root or anything else. Please ensure that the horizontal axes have the same extension so the plots are comparable, and that the whiskers are not cut off by the figure bounding box.

- Table 7: please indicate in the table caption what "their counterparts" are. The entire table is unclear to me; how do the top and the bottom half differ? What does "Model with 5 factors and category dummy variables" as a caption to the bottom half (or is it?) refer to, in contrast to the top half? The text says on p. 23 that "the horizon is one to eight-week ahead", but the table says "Horizon = 8" - please clarify whether the horizon is eight weeks or *up to* eight weeks.

- References: please provide full details for Loeb (2015)

Typos:

- p. 3, 2nd para, l. 3: "The model which is subject to structural break" - either add "a" or change to "breaks"

- p. 3, 2nd para, l. 7: remove "in" after "including"
- p. 4, l. 2: "mention" -> "mentioned"
- p. 4, l. 3: "values" -> "value"
- p. 4, 2nd para, l. 2: "the change of" -> "changes in"
- p. 5, l. 5: "forecast" -> "forecasts"
- p. 6, 2nd para, l. 3: "structure" -> "structural"
- p. 6, third equation block: italic and upright versions of β , X and Q are mixed, which is painful to me (also on p. 7)). Please ensure proper mathematical typesetting. Lowercase x_{T+h} and uppercase X_{T+h} is used inconsistently here.
- p. 7, equation for the bias correction: lowercase ω is used both as the starting index of the summation (with the ending index indicated by W - why mix Greek and Latin?) and as the summation index itself.
- p. 7: "The estimated bias are" should be "is"
- p. 10, equation: can be slightly simplified by removing " $=\eta, \eta$ "
- p. 19/20: Figure 3(c) is not "in the bottom-right corner", but in the bottom left.

Reviewer #2: This is an interesting paper that is trying to investigate the forecasting performance of several ADL models over retailer product sales. The study is particularly focusing on the effects of structural breaks originating from marketing activities over the products.

Main Comments:

1. I found the structure of this paper very confusing. For example, the introduction and literature review section are very poorly written with many overlaps and repetitions that are not at all informative for the reader. The contribution of the paper, as outlined, is very weak.
2. In the introduction the authors should clearly indicate what is the model they are introducing and why. On top of that, it should be made clear to the reader why the specific models seem appropriate for the retail forecasting exercise. It is quite astonishing that the reader does not get a gist of what he/she is going to see unless he/she reaches page 10.
3. The literature review is very short and several references are packed all together without any meaningful commentary (for example in page 5). I would really be interested to see what are the findings of studies assuming constant marketing activities, as this would highlight/clarify/validate potentially comparisons with the models at hand.
4. The text is characterized by some generalizations that make the reader confused on what the authors are claiming. For example, in page 3 '... The data in retailer product sales... macroeconomics).' I would suggest a thorough read-through to the authors in order to make the text more to the point.
5. Section 3 and 4 seem a bit redundant the way they are presented. I would expect to see a methodology section, where these two sections could motivate/inform the selection of models in section 6. The ideal approach would be that section 3 and 4 are reduced substantially and included in the commentary of section 6 or if needed in an appendix.
6. In section 5, it should be explained what the display and feature percentage is along with the motivation of the selection of these inputs. In my view, the data section should have been after the intro/literature review.
7. In terms of the analysis during structural breaks, it is interesting that the authors do not make explicitly clear how they detect, test or analyse their results for structural breaks. It would be expected that in that type of paper where forecasting performance is evaluated within the presence of structural breaks that vast emphasis would be given on that issue. The general information provided in section 4 are not enough in my opinion. What tests have been done and what was the result? Except from the Chow test, have the authors investigated the Andrews approach (2003, Econometrica) or Fixed Regressor Bootstrap? The vague analysis on that aspect is a main shortcoming of this paper.
8. In the modelling part, I would expect to see a small description at least of the benchmark model, the base-lift model. Also, it is not well explained why Lasso is used two times. What motivates the authors to double apply Lasso shrinkage? Is there any similar study suggesting that? What are the expected benefits of this approach? Given that this paper comes down to a horse-racing application between ADL models constrained by Lasso operators, it is important to clarify why the ADL-raw models is combined with the ADL-own model? Wouldn't be a logical question from the reader as to why the Lasso operator is not used in a third stage, for example? A lot of discussion is provided before-hand (in cases unnecessary), but at section 6 that the reader wants to understand the reasoning and process of modelling, he/she struggles to do so because of the lack of explanations. Another important shortcoming for this work.
9. In the experimental design section, the representation of the competing models is not clear. The authors should have included perhaps a table or a clear text on what is the differences between models in a concise and transparent way. Additionally, I am not sure why the roll-forward period is equal to two weeks, while the forecasts are done for 1, 4 and 8 weeks. The authors should explain why they used sMAPE over MAPE or why MASE is more informative. In terms of statistical accuracy, RMSE and Theil-U statistics are also quite often

used. One final point is why are the authors inclined to use the adjustment of Cooper et al. (2009). This was not quite clear to me.

10. The section 8 is quite extensive in terms of results in tables, but the comments are quite limited and badly written. For example, is there consistency in ranking across statistical measures? What exactly is the importance of Table 3? At such a forecasting exercise, a Diebold-Mariano test is needed to assess the significance of the differences in performances. The Wilcoxon Sign Rank is not enough. It is also not quite clear what Table 4 is offering. In Figure 3, is there a particular reason why these six product categories are selected for analysis?

11. Table 5 should probably be within the analysis of the determinants of the forecasting performance. I also find the explanation of the process confusing. Why is this factor selection followed? Have the authors explored another factor analysis?

Minor comments:

1. The writing of the whole paper is not very good. A lot of proof-reading is required. Some examples:

* 'Under such a circumstance...activities' in the abstract.

* '...the generated forecasts may potentially...' in the abstract

* SKU abbreviation should be defined in the abstract and in text and then explained perhaps on a footnote.

* '...proposed holistic methods to generate...' page 4, line 57

* Footnote 7 should refer to figure 1.

2. Sometimes writing lack of academic standard. For example, there is no numbering in equations, there is extensive use of bullet points within text, tables are not explained with footnotes, many equations appear within text, the tables style is not uniform etc.

3. The paper is not well positioned in the OR forecasting literature.

Overall, based on the above I am inclined to reject the paper due to its lack of solid contribution, convincing results, presentation and academic rigour.

Close