Reaction to referee report Nr. 2

Sven-Kristjan Bormann

> 1) the use of factor analysis for extracting latent dimensions from a correlation structure is a good idea, but: a) as the author specifies in a footnote (on page 19), it's an exploratory method, i.e. its results can't be generalized to the population, so using them to confirm (or reject) a priori knowledge is inappropriate, and b) in this case, results of factor analysis are obvious given the correlation analysis in table 3, that already shows the presence in the data of two latent factors strictly related to short and long term sentiment respectively.

I agree that the results of the factor analysis are kind of obvious given the results from the correlation analysis. However the interpretation of the correlations and the factor analysis is driven by the proposed theoretical properties of sentiment data. Hence the factor analysis helps to see these properties more clearly than by just looking at the correlation matrix. The factor analysis acts here as a tool to expose the theoretical predictions differently and to lay the ground for further analyses.

"2) Several times the author justifies his methodological choices with unsuitable arguments, such as the lack of software resources for running appropriate testing."

I agree that lack of software resources might look like an inappropriate excuse for certain methodology choices. However I made these clarifications only to show that I am aware of certain problems with performing certain tests, but that I was not able to solve these problems. Given that I saw some potential problems in sections of the paper which are not my main contributions, I can leave out these justifications. Besides this specific point, I am not aware of any further unsuitable justifications of my methodological choices. If they are nevertheless still present, I will correct them.

> 3) as the author himself specifies, his results do not provide useful information to investors using sentiment data. Maybe the analysis

> might lead to more interesting results if a confirmatory approach is used, such as a latent variables model, and if new variables are included in the analysis, as well as relationships between factors.

It should be feasible to make a confirmatory analysis on a different dataset. The analysis would take the results of the already presented exploratory factor analysis and the psychological definitions of sentiments as input for the model to be constructed and confirmed. However I am sceptical whether the confirmatory analysis would provide any more useful information for investors than the exploratory one besides maybe a confirmation of the previous results. Both analyses – confirmatory and exploratory – rely for the same data on the same information from the same covariance matrix. The only real differences between to the two analyses are the restrictions on the covariance matrix and the goals of the analyses. If I added additional variables and estimated a latent variables model then I would end up probably with a new model, which might require adding another theory. I accounted already for the relationships between the factor through an oblique rotation in the factor analysis. Therefore I doubt to get any new relevant results from this exercise.

Suggested changes:
"1. the author should give a more detailed description of data;"
In my opinion I give a sufficient description of the data by describing in section 3 how the participants are selected and how the indices are calculated. A link to the sample survey is also given. Descriptive statistics about the data are presented in section 4.1. If further details are desired, I will add them to provide a clearer description of the data.

"2. I suggest to add the significance levels for correlation coefficient in table 2"
I will add them and thank the referee the suggestion.

"3. the author should specify why do tests results in Appendix A refer to 1-99, 100-300, 301-415 periods;"
I will add the specification and justification of the periods to the appendix.

"4. in tables 9, 10 and 11 in Appendix A the last two cells in the last row must shift on the right (I suppose they are confidence interval limits);"
I will correct the formatting issues.

"5. on pag. 13, the author says: "The value of partial autocorrelation is usually two...", does he mean the order of partial autocorrelation?"

Yes I meant “order” instead of “value”. Thanks for indicating this mistake.

“6. results of t-tests and variance ratio tests in Appendix A could be reorganized in order to avoid the duplication of all descriptive statistics;”
I will find a way to do it.

“7. the normality analysis (page 21) is not clear. Also, it leads to trivial conclusions, e.g. the (well known) asymmetry of stock market indices.”
The normality analysis is about the sentiment indices and *not* about the stock market indices. I will clarify this part to avoid further confusion.