Dear Reviewer 2,

Thank you for your comments. Before answering your specific remarks, we would like to refer to your general comments. As we were requested to provide the response first, unfortunately, you cannot find the results of our work in the paper yet. However, we are committed to implement all the below mentioned comments in the improved version of the paper.

You claim that: "The paper is poorly written, not very well motivated and, there are several typos all around the text". Of course, current version of the text is just a working paper, which will be improved in the subsequent stages. At this point, we can ascertain that our intention is to revise the flow of the text in order to improve its readability but especially our intention before the text is finally accepted is to provide additional proofediting in order to ensure the highest quality of English used.

Another general issue that you raise refers to the motivation. You claim that "I find the motivation of "atheoretical" models vs. theoretical models redundant as there are not any theoretical models competing with the models in the paper." Although you are right that there are no theoretical models used in the paper, our intention was just to motivate the path pursued for obtaining forecasts. We wanted to provide justification, although a limited one, as it is not the main issue in the paper, why we think that such forecasts might be of value and that one of the avenues, which can be pursued in the research and was adopted by us, is to rely on results of "atheoretical" models.

Of course, we are willing to benefit of your expertise during the review process on eliminating the flaws associated with lack of justification for "ad hoc decisions during the execution of the methodology". Our intention was not to leave the reader looking for justifications but to provide a concise overview of scientific developments of our methodology. It is good that you mention it at this stage, when such flaws can still be amended. Our judgement might have been a bit blurred by our long research in this field and some points important for the reader might have been unintentionally missed. It can be improved in the subsequent versions of the article.

 Please find below our responses to your specific questions and doubts

1. On the motivation of "atheoretical" models vs. theoretical models: Many of the models in the paper can be thought as reduced form models derived from structural models. Perhaps more importantly, many of the seminal papers on this issue such as Sims (1980) are not cited, which makes the discussion very loosely grounded.

Ad. 1. In fact, the convention used in the article is in the frame of dynamic simultaneous equations models. Although it may seem that the structure of the model is arbitrary chosen, it was substantiated by our previous research (Białowolski, Kuszewski, & Witkowski, 2010). We did not recall Sims (1980) because we found that we did not completely follow his postulates. We would like to recall that his main points were that:

(1) There is no a-priori distinction between exogeneous and endogeneous variables, i.e. there are no causal relations built between categories describing behaviour of an economic system;

(2) No constraint on the parameter values in the model are imposed, in particular, it is not assumed that certain parameters are equal to zero, which leads to elimination of variables associated with these parameters in the final form of the model;

(3) There is no search performed for an underlying economic theory, which could be primary with resepct to the model.

Although we are completely in line with the postulate (3) our approach does not comply with (1), while we define endogeneous GDP, CPI and UNE. However, considering your remark, we are of course willing to include the reference and briefly discuss its implications for our modelling approach, especially as we have used diversified approach with respect to postulate (2) in dynamic factor approach and Bayesian averaging.

2. Closely related to the first bullet, I think the motivation should be rather on the forecasting methodology, i.e. model averaging vs. factor models and the use of tendency survey data.

Ad. 2. Our approach is in fact mostly motivated by an ability to generate the most accurate forecasts. We were convinced that this message was clearly stated in the text. Obviously, we can emphasize this point in an amended version of the text. Yet it must be emphasized that the spirit of the paper is to use a few methods of estimation (factor models, a few types of Bayesian averaged models), however, the final choice needs to be based on "something". We trust that considering the target of the paper (which is to develop a forecasting tool) the forecasting performance is of use as a criterion of choice.

3. The term "economic situation indicators" can be replaced with another term potentially "coincident" or "leading" indicators.

Ad. 3. In search for the best terminology, we can of course agree to using "coincident" and "leading" indicators (both groups are present in the text) instead of "economic situation indicators".

4. The ordering in equations (2) is key for the paper as it as followed throughout the paper and therefore, it requires more discussion. It would be also nice to see the forecasting results using, for example, an alternative ordering that yields the best results compared to other remaining alternatives.

Ad. 4. As already mentioned, in our previous work (Białowolski et al., 2010) we have been exploring all six possible orders within the endogenous variable vector {GDP, UNE, CPI}, {GDP, CPI, UNE}, (CPI, GDP, UNE}, {CPI, UNE, GDP}, {UNE, GDP, CPI}, {UNE, CPI, GDP}. We are also not in a position to interpret the order of variables. Our conclusion is solely that for quarterly data of years 1990-2012 the model described by the first permutation of endogenous variables (out of the listed above) yielded the most accurate forecasts. Consequently, based on the adopted methodological approach, the model became our choice for subsequent analyses. We would like to stick to the approach in this respect as

(1) We cite the analyses that lead to such a conclusion and the selected order was superior with respect to forecasting all three macroindicators of interest. Of course it is possible that in the future revision might be necessary but we are in a position that the time has not come yet.

(2) We do not want to present all the permutations as it would lead to considerable increase of the word count of the paper making it incomprehensible.

5. Regarding to equations (2) which particular variables are subject to endogeneity: If these are the GDP, UNE and CPI, then the question is on the efficiency of the use of equation-by-equation estimation. As the error terms are correlated and, as far as it can be understood from the exposition of equations (2), the system seems not to be just identified, an efficient way of estimation would be system estimation. Finally, if the authors would want to impose a specific ordering then why not using structural VAR instead these equations?

Ad. 5. The main problem behind the use of systems, especially the VAR models, is the number of degrees of freedom. The sample size is very low compared to the number of parameters of interest. Both the system estimator and especially any kind of VARs share one problem: we would need to estimate the set of parameters at once, which is impossible due to the negative number of degrees of freedom in such a case. Applying Bayesian averaging allows for estimating a number of equations with just a subset of all the considered regressors in each. Afterwards the estimates are averaged. In the frequency approach we follow similar procedure, however drop the unnecessary variables and re-estimate the equations. In the final step, however, the number if variables in the latter case is much lower than initially. Of course if we had a big number of observations, we would share the Rewiever's opinion, still the time series does not provide us this comfort due to its length. Additionally, we did not use neither structural nor non-structural VAR to obtain our forecasting goals because our main goal was the accuracy of forecasts. Application of VAR models requires discussion of imposed constraints, which in turn is associated with modeller's intervention. Our approach is to limit this intervention in order to facilitate the forecasting process and reproduction of forecasts on quarter-to-quarter basis.

6. It would be nice to include a brief outline of details of Bayesian model averaging.

Ad. 6. This can of course be done – we referred to other papers (including our own) in order not to be redundant, yet surely we can add the description to the article.

7. On page 7, when introducing the methods of averaging, the third method that takes collinearity into account is surprising. It should be better motivated maybe at earlier stages why the authors also use this method.

Ad. 7. Naturally, this also can be done. We found it quite natural to compare the performance of „pure" and „collinearity-modified" frequency model averaging: contrary to the „classical" algorithm, in the frequency approach the final model is estimated again and the whole Bayesian approach is used just to select the variables for the final step. That makes the estimation of just a single equation

(a set of three equations in this case) sensitive to collinearity. Thus dropping the collinear variables seems an idea worth considering. We can obviously expand on this rationale in the paper.

8. On page 8 (we assume that you are referring to page 11 because on page 8 there are no indicators), I think some of the indicators are not used at all. This should be indicated explicitly with the reasoning behind why it was dropped. There should be a sensitivity check on the potential effects of these ad hoc choices.

Ad. 8. Estimation of our models was performed on 68 quarters of data and the set of variables comprised 56 indicators. Thus leaving only 12 degrees of freedom and even lower number if lagging values of the factor were included. With such a low number of degrees of freedom the results were volatile and in some situations it was not possible to obtain any dynamic solution. It was proposed, following approaches used in factor models based on micro data, to exclude from the factor indicators the ones that produced non salient coefficients (Brown, 2006; Matsunaga, 2011; Osborne & Costello, 2004). Using indicators that are not salient does not add any value, as such indicators are usually only source of a noise a do not contribute to the variation of factor. In such a case, a non-salient indicator has very low correlation with the factor, which leads to a conclusion that existence of a common factor driving the data might be questioned but also being strong indication for choosing different set of indicators (Athanasoglou, Weziak-Bialowolska, & Saisana, 2014; Saisana & Weziak-Bialowolska, 2013). Hence, we decided to base our solution on the initial set of indicators that were salient. In order to obtain the set of salient indicators we used preliminary solution and extracted factor loadings with standard PCA. Of course, we are willing to perform a sensitivity analysis to the solution depending on the choice of "saliency|", i.e. the threshold at which indicators are dropped from the analysis. Such analysis would probably increase the appeal of the paper.

9. On page 8 (probably again you mean page 11 and we refer to it), the table-ish display of the indicators should be replaced with a nice table or should be moved totally to the appendix with some explanations in the main text.

Ad. 9. We agree. It will be done in the revised version of the article.

10. My impression from the information in the text is that the factors are estimated using the principal components rather than likelihood based inference. In this sense, what is the use of dynamic evolution of the factors displayed at the bottom of page 10 in the forecasting exercise.

Ad. 10. No, it is not true. Only for the preliminary selection of indicators (see answer to 8) PCA has been applied. After the set of indicators was chosen likelihood based estimation was performed.

11. Table 1 should be replaced with more informative table including not the mnemonics but the description of the variables.

Ad. 11. We of course intend to do it in the modified version of the paper. Our initial justification was to limit the space devoted to presentation of various specifications. We also neither track nor interpret the interrelation between variables and the factors (after we select the set), so we did not want to include the information, which is subsequently not discussed. However, we feel that, as you raise the issue, it should be accounted for in the revised version of the paper.

12. Why not extracting the factors from the complete set of indicators rather than splitting the information set into three parts? In this case what is the forecasting performance of these models. It should be provided a sensitive check on this choice.

Ad. 12. We did not exactly split the initial set into three sets. We have applied PCA to extract three factors and searched for such indicators to those factors that will be salient in a static solution (see answer to 8). Depending on the choice of the threshold for inclusion of the indicators to factors 1-3 we would obtain different wider or narrower set of indicators for each of the factors. As mentioned also in 8, we are eager to make a sensitivity check in the revised version of the paper.

We are looking forward to hearing your comments and further suggestions to work on the improved version of the paper.

Best regards,
Authors

References

Athanasoglou, S., Weziak-Bialowolska, D., & Saisana, M. (2014). *Environmental Performance Index 2014 JRC Analysis and Recommendations - JRC Science and Policy Reports* (Vol. Report EUR). Luxembourg: European Union. http://doi.org/10.2788/64170

Białowolski, P., Kuszewski, T., & Witkowski, B. (2010). Business survey data in forecasting macroeconomic indicators with combined forecasts. *Contemporary Economics*, *4*(4), 41–58. Retrieved from http://we.vizja.pl/en/issues/volume/4/issue/4#art178

Brown, T. A. (2006). *Confirmatory Factor Analysis for Applied Research*. New York, NY: The Guilford Press.

Matsunaga, M. (2011). How to Factor-Analyze Your Data Right : Do's, Don'ts, and How-To's. *International Journal of Psychological Research*, *3*(1), 97–110.

Osborne, J. W., & Costello, A. B. (2004). Sample size and subject to item ratio in principal components analysis. *Practical Assessment, Research & Evaluation*, *9*(11).

Saisana, M., & Weziak-Bialowolska, D. (2013). JRC Statistical Audit on the Environment and Gender Index. In IUCN (Ed.), *The Environment and Gender Index (EGI). 2013 Pilot*. Washington D.C.