March 30, 2020

Dear Professor Venetis,

I am writing you regarding your manuscript, “A replication of ‘A Quasi-Maximum Likelihood
Approach for Large, Approximate Dynamic Factor Models’ (Review of Economics and Statistics, 2012)” (*Manuscript Number 3413; Discussion Paper Number No. 2020-5*).

I have now received reports from one anonymous reviewer, an author of the original paper, and a co-authored comment that has been subsequently published as an *Economics E-Journal* discussion paper. Based on these comments, I am inviting you to submit a revision of your manuscript.

I would like your revision to address the following comments by the reviewers:

From the anonymous reviewer:

1. “An important difference between your dgp and Doz et al (2012) is that in eq (1) you allow for lags. It is an important contribution that you're making and I think you should stress this more in section 2.”

2. “A measure which is equally important to report is the MSE between the estimated common component and the simulated one.”

3. “The static representation of (1)-(2) is valid of course but you cannot hope to estimate the dynamic factors f once you have the static ones F. In fact what you say is that you are interested in F, not in f. But that makes sense to me only if the common component is your interest but if you want to attach a "meaning" to the factors then f should be your interest and you cannot recover them from F. For this reason I think studying the MSE of the estimated common component instead would be important since it is not affected by identification issues.”

4. “On page 8 step 3, why do you extract qs PC? shouldn't be q(s+1) PC?”

5. “There are two recent papers studying the asymptotic properties of the EM estimation of stationary factor models with singular factors (Barigozzi Luciani, 2019a) and of non-stationary dynamic factor models, where factors are loaded with lags, s>0 (Barigozzi Luciani, 2019b) please refer also to those papers and the simulation results therein.”

From the original author:

6. “The authors propose to compare 5 estimations methods, which are listed page 8. Some remarks must be done before going further :
   – Method 3 should be applied using q(s+1) principal components and not qs since Ft has dimension q(s + 1). This might be a typo but this error also appears in the second and fourth paragraphs of page 10.
   – Method 4 is irrelevant : why should we extract 1 principal component when we know that the static factor has dimension q(s + 1) = 4?

– Method 5 is irrelevant too : why should we extract 1 principal component and take its first lag when we know that the dynamic factor has dimension q = 2 and should indeed appear with one lag?"

7. "The authors are aware of the two last problems and say (page 10) that it is a way to see how the PC estimators behave under misspecification. But this section aims at comparing QML estimators with other estimators. Thus, if the dimension of the factors is well specified for the QML estimator, it should be also well-specified for any PC estimator which is considered. The behavior of any of those estimation methods under misspecification could also be considered, but it is another topic."

8. "Methods 1 and 2 are indeed an extension of DGR who did not consider the case where the factors enter the measurement equation with lags. As intialization is always an issue, when using numerical optimization methods, I think the authors should give more details on the way they proceed."

9. "In particular, for method 2, it is unclear how $\hat{f}_t$ is estimated using the two-step method of Doz et al (2011). Indeed in the first step of in Doz et al (2011) method, the factors are estimated by PCA, but the model is a static one (no lag in the measurement equation), which is not the case here. In particular, in the present case, using a PCA to extract only q factors in the first step would certainly not be the right way to proceed. If it is what has been done, it could explain the disappointing results obtained for TS."

10. "However in the extension part : the number of factors in methods 4 and 5 (and possibly in method 3 if q s instead of q(s + 1) is not a typo) should be modified and the initialization procedures used in methods 1 and 2 should be made more precise."

From Poncela and Ruiz's discussion paper.
*NOTE: I have added some clarifying notes below some of the points to provide direction about how you should handle these points.*

11. "In the context of the D-DFM there is an even more important identification issue related with identifying simultaneously the lag order of the VAR model for the factors, p, and the number of lags, s, in equation (1)."

12. "Although LV20 conclude that a correct specification of the underlying dynamics is of paramount importance, they do not challenge the KFS factor extraction when the relevant quantities, p, q and s, are unknown."

13. "This example just illustrates the difficulty involved in the identification, even in the context of this very simple S-DFM, when the idiosyncratic errors are temporal and cross-sectionally correlated. Identification in the context of the D-DFM is even more challenging. This is an issue that deserves further research."
    *NOTE: Comments [11]-[13] regarding identification are very important and should be prominently discussed in your paper.*

14. "LV20 assume them [the errors] to be standard normal. Although normality could be a good approximation to start with, it could also be useful to investigate the performance of KFS under other distributions of the factors and/or idiosyncratic components; see Barigozzi and Luciani (2019) who carry out Monte Carlo experiments assuming that the

innovations are Student-t. Our guess is that, while point estimates of the factors may not be severely affected by non-normal errors, interval estimates could be."
*NOTE: You do not have to undertake more MC experiments with Student-t errors, but you should discuss this point in your revision.*

15. "What is the point in changing the dynamics of the idiosyncratic components and, consequently, the definition of the matrix T with respect to that in Doz, Giannone and Reichlin (2012) by allowing each idiosyncratic noise to have their own autoregressive parameter if then the results are reported only for the case in which all parameters are the same?"
*NOTE: This needs either to be explained better, or you should conduct additional MC experiments with different autoregressive parameters.*

16. "LV20 claim in the abstract that the persistence characteristics of the observable series play a crucial role, they only report results for a very persistent factor, with its autoregressive parameter being 0.9. In order to analyse the role of persistence on the results, one should at least consider different levels of persistency in the factors. This is indeed a very crucial question for users of DFMs."
*NOTE: You should conduct additional MC experiments to illustrate the importance of persistence.*

17. "For the factors to be pervasive, a large enough number of weights should be different from zero. Simulating the weights from a N(0; 1), a large number of simulated weights could be close to zero and, consequently, it could generate weak factors…"
*NOTE: You do not have to conduct additional MC experiments with weights from the uniform, but this point should be discussed.*

18. "…in their Monte Carlo experiments, LV20 drop the replicates for which TR < 0.05. The authors are throwing away replicates for which the method is not working. Obviously, the results reported are better than they should be. It is important to know at least how many replicates are discarded."
*NOTE: At the very least, report how many replicates were discarded.*

19. "…in the case of the D-DFM considered as DGP, it is not clear what the authors are estimating when implementing the estimation procedures proposed by Doz, Giannone and Reichlin (2011, 2012) that are designed for S-DFMs."
*NOTE: Clarify what you are doing in your implementation.*

20. "Seventh, in any case, our main concern is about the design of the Monte Carlo experiments being too restrictive as to be of real interest for users in the sense that the number of factors, q, and the lags, s and p, are assumed to be known. For the Monte Carlo experiments to be of real interest for users, one should consider uncertainty about the number of factors and their dependences."
*NOTE: Your revision should acknowledge that the assumptions that (i) the number of factors are known, and (ii) the lags s and p are known, are unrealistic assumptions that limit the usefulness of your Monte Carlo experiments for applied work. Then stress that these are important issues for future research.*

A revised version of your manuscript that addresses the instructions above will be reconsidered for publication. If you choose to resubmit a revision, be sure to include a point-by-point response to the items listed above.

Once again, thank you for submitting your manuscript to *Economics; The Open-Access, Open-Assessment E-Journal*. I look forward to receiving your revision.

Sincerely,

Bob Reed
Co-Editor, *Economics E-Journal*