We congratulate the authors for this insightful paper. Here we suggest a method to address two concerns:

1. The potential underestimation of the number of factors $K$;

2. The potential non-sparseness of the estimated principal orthogonal complement.

Point 1 is addressed by using a larger $K$. With pervasive factors assumed in the paper, it is relatively easy to find such $K$. However, in an analysis of macroeconomic data for example, there can be a mix of pervasive factors and many weaker ones; see [3], [5] and [4] for a general definition of weak factors.

In [6], a monthly data of $p = 132$ U.S. macroeconomic time series from 1959 to 2003 ($n=526$) is analyzed. Using principal component analysis (PCA) [2], the method in [5] and a modified version called the autocovariance-based factor modeling (AFM) (details omitted), we compute the average forecast errors of 30 monthly forecasts using a vector autoregressive model VAR(3) on the estimated factors from these methods with different number of factors $r$ (Figure 1). While 3 pervasive factors decrease forecast errors sharply, including more factors, up to $r = 35$, decrease forecast errors slower, showing the existence of many “weaker” factors.

Hence it is not always possible to have “enough” factors for accurate thresholding of the principal orthogonal complement, which can still include contribution from many weak factors and is not sparse. Points 1 and 2 can therefore be closely related, and can be addressed if we regularize the condition number of the orthogonal complement instead of thresholding. While [7] restrict the extreme eigenvalues with a tuning parameter to be chosen, we use the idea of [1] (properties are not investigated enough unfortunately). I and my PhD student Charlie Hu are studying its theoretical properties.
We simulate 100 times from the panel regression model

\[ y_t = X_t \beta + \epsilon_t, \quad \beta = (-0.5, 0.5, 0.3, -0.6)^T, \]

with \( x_{it} \) being independent AR(1) processes and \( \epsilon_t \) the standardized macroeconomic data in [6] plus independent \( N(0, 0.2) \) noise. Following Example 5 of the paper, we estimate \( \Sigma_{\epsilon}^{-1} \) using different methods and plot the sum of absolute bias for estimating \( \beta \) using generalized least square (GLS) against the number of factors \( r \) used in Figure 2. Clearly regularizing on condition number leads to stabler estimators.

---

**Figure 1:** Average forecast errors for different number of factors \( r \).

**Figure 2:** Sum of absolute bias (averaged over 100 simulations) for estimating \( \beta \) using GLS against the number of factors \( r \) used in POET (\( C=0.5 \)) and the condition number regularized estimator. Bias for least square method is constant throughout.
Table 1: Comparisons of the risks of portfolios by using POET ($C = 0.5$ and condition number regularized estimator).

<table>
<thead>
<tr>
<th>$K$</th>
<th>Proportion of time POET outperforms</th>
<th>Percentage of average risk improvements</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.40</td>
<td>-4.07%</td>
</tr>
<tr>
<td>2</td>
<td>0.46</td>
<td>-2.50%</td>
</tr>
<tr>
<td>3</td>
<td>0.56</td>
<td>0.66%</td>
</tr>
<tr>
<td>4</td>
<td>0.56</td>
<td>0.71%</td>
</tr>
</tbody>
</table>

Parallel to section 7.2, we compare the risk of portfolios created using POET and the method above. Again Figure 3 and Table 1 show stabler performance of regularization on condition number.

Figure 3: Risk of portfolios created with POET ($C = 0.5$) and condition number regularized estimator.

References


