Une évaluation du test de suffisance de la prochaine valeur propre (NEST)

Pier-Olivier Caron[1]

[1]Université TÉLUQ

Présentation au 45e congrès de la SQRP

Dans le cadre du Symposium « Développements en analyse quantitative et enseignement de la statistique »

le 26 mai 2023

Préliminaire

Déterminer les nombres de facteurs à retenir dans une analyse factorielle exploratoire est l'un des problèmes méthodologiques toujours ouverts.

Une pléthore de techniques existent pour répondre à cette question. De nouvelles sont publiées chaque année.

L'une des techniques prometteuses est le test de suffisance de la prochaine valeur propre (NEST; Achim (2017)) qui montre un excellent rendement Brandenburg & Papenberg (sous-presse).

Elle n'a toutefois pas été systématiquement comparée à ses concurrentes.

Objectif

Comparer le rendement de différentes techniques recommandées (Auerswald & Moshagen, 2019) en plus de NEST.

- NEST (Achim, 2017)
- l'analyse parallèle (Horn, 1965)
- la corrélation partielle moyenne minimum (Velicer, 1976)
- le test de χ^2 séquentiel (Lawley, 1940)
- la méthode de Hull (Lorenzo-Seva et al., 2011)
- le critère de Kaiser empirique (Braeken & Assen, 2017)

Techniques

Parallel analysis (PA)

Rééchantillonner les valeurs propres d'un jeu de données sans aucun facteur (aucune corrélation entre les variables) ayant les mêmes caractéristiques que le jeu de données cible (mêmes nombres de variables et de participants).

On conserve les premières valeurs propres empiriques supérieures à celles simulées.

Next Eigenvalue Sufficiency Test (NEST)

Tient compte de l'erreur d'échantillonnage, comme l'analyse parallèle, mais aussi de la séquence des tests.

Le test utilise une matrice de corrélation contenant les k dimensions déterminées auparavant.

Lorsque $k = 0$, le test est équivalent à l'analyse parallèle.

On conserve la k^e dimension si la valeur propre est supérieur à celles simulées.

Empirical Kaiser criterion (EKC)

La distribution des valeurs propres suit asymptotiquement suit une distribution de Marchenko-Pastur.

$\lambda_0 = (1 + \sqrt{p/n})^2$

pour la 1^{ere} et corrigée pour les suivantes

$\lambda_j = \max\left(\frac{p \sum_{i=0}^{j} \lambda}{p-j-1}\left(1 + \sqrt{p/n}\right)^2, 1\right)$

La valeur de 1 est le minimum (comme le critère de Kaiser).

On conserve les premières valeurs propres empiriques supérieures aux critères.

La méthode de Hull (HULL)

Semblable aux variantes non graphiques de Cattell, la méthode de Hull tente de trouver un coude dans les valeurs propres.

Au lieu d'utiliser les valeurs propres par rapport au nombre de facteurs, la méthode de Hull s'appuie sur des indices de qualité d'ajustement (CFI) relativement aux degrés de liberté du modèle proposé.

On conserve le dernier CFI sans améliorations.

Minimum average partial correlation (MAP)

Calculer en éliminant successivement la matrice de loadings des composantes des premières premières composantes la matrice de corrélation originale pour ensuite obtenir la matrice de covariance partielle.

On conserve le nombre de facteurs associé à la moyenne des corrélations partielle la plus faible.

Sequential χ^2 model tests (SMT)

Un test séquentiel de l'estimation du maximum de vraisemblance (ML) selon lequel la matrice de covariance du modèle est égale à la matrice de covariance.

On conserve la première structure dont le χ^2 est non-significatif.

Méthode

Simulation

Une simulation avec des structures factorielles synthétiques (Caron, 2016, 2019).

Les structures sont basées sur

- 24 variables;
- de 1 à 8 facteurs;
- des loadings variant entre .40 et .80;
- des corrélations interfactorielle de .00 à .30;
- trois tailles d'échantillons, 120, 240 et 480;

En tout, ce sont 360 scénarios qui sont répétés 1000 fois.

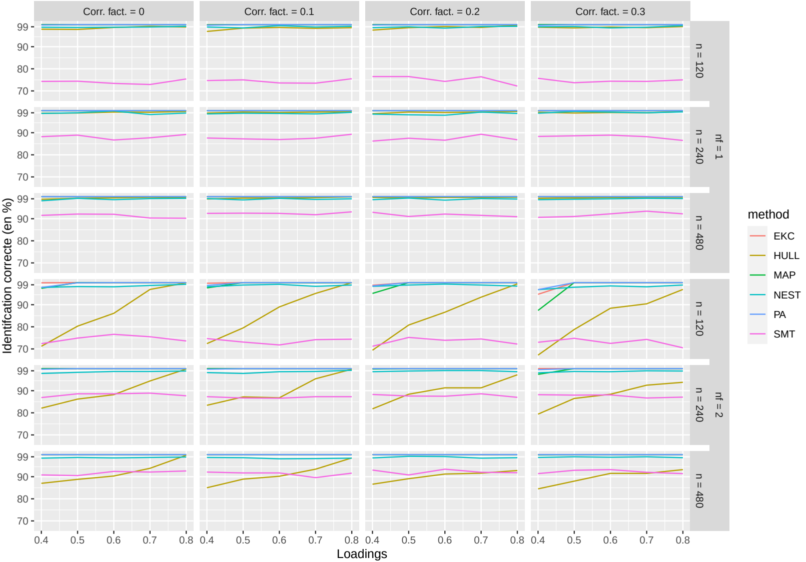
Rendement

Le rendement est évalué en termes de

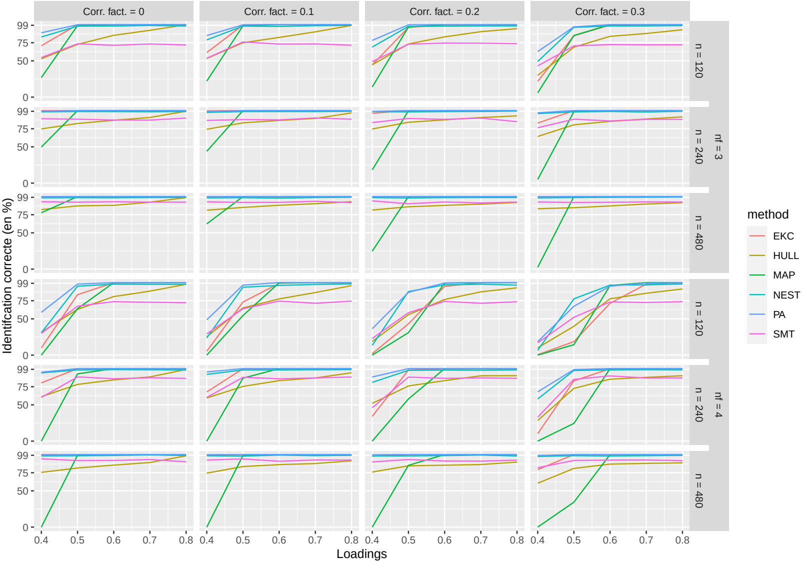
- précision (identification correcte de la dimensionalité);
- biais (tendance à sur- ou sous-estimer la dimensionalité).

Résultats

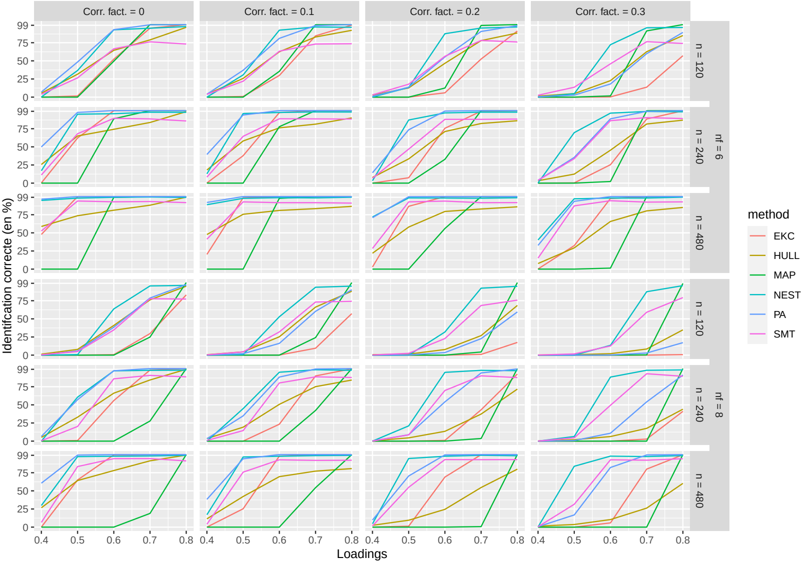
Cas très faciles



Cas faciles et intermédiaires



Cas difficiles et très difficiles



Limites

De la simulation

- Des structures factorielles avec des loadings uniformes.
- Des valeurs propres égales pour tous les facteurs.

De NEST (et toutes les techniques basées sur les valeurs prores, comme PA et EKC)

- Souffre du paralogime de l'affirmation du conséquent

Conclusions

Si la plupart des techniques réussissent bien les scénarios faciles, NEST se distingue particulièrement dans les difficiles.

Tester les techniques avec des structures factorielles plus réalistes et plus variées.

Rester vigilant.e.s!

Références I

Achim, A. (2017). Testing the number of required dimensions in exploratory factor analysis. *The Quantitative Methods for Psychology*, *13*(1), 64–74. https://doi.org/10.20982/tqmp.13.1.p064

Achim, A. (2020). Esprit et enjeux de l'analyse factorielle exploratoire. *The Quantitative Methods for Psychology*, *16*(4), 213–247. https://doi.org/10.20982/tqmp.16.4.p213

Achim, A. (2021). Determining the number of factors using parallel analysis and its recent variants: Comment on lim and jahng (2019). *Psychological Methods*, *26*(1), 69–73. https://doi.org/10.1037/met0000269

Références II

Auerswald, M., & Moshagen, M. (2019). How to determine the number of factors to retain in exploratory factor analysis: A comparison of extraction methods under realistic conditions. *Psychological Methods*, *24*(4), 468–491. https://doi.org/10.1037/met0000200

Braeken, J., & Assen, M. A. L. M. van. (2017). An empirical kaiser criterion. *Psychological Methods*, *22*(3), 450–466. https://doi.org/10.1037/met0000074

Brandenburg, N., & Papenberg, M. (sous-presse). Reassessment of innovative methods to determine the number of factors: A simulation-based comparison of exploratory graph analysis and next eigenvalue sufficiency test. *Psychological Methods*. https://doi.org/10.1037/met0000527

Références III

Caron, P.-O. (2016). A monte carlo examination of the broken-stick distribution to identify components to retain in principal component analysis. *Journal of Statistical Computation and Simulation*, *86*(12), 2405–2410. https://doi.org/10.1080/00949655.2015.1112390

Caron, P.-O. (2019). Minimum average partial correlation and parallel analysis : The influence of oblique structures. *Communications in Statistics - Simulation and Computation*, *48*(7), 2110–2117. https://doi.org/10.1080/03610918.2018.1433843

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, *30*(2), 179–185. https://doi.org/10.1007/BF02289447

Références IV

Lawley, D. N. (1940). The estimation of factor loadings by the method of maximum likelihood. *Proceedings of the Royal Society of Edinburgh*, *60*(1), 64–82. https://doi.org/10.1017/S037016460002006X

Lorenzo-Seva, U., Timmerman, M. E., & Kiers, H. A. L. (2011). The hull method for selecting the number of common factors. *Multivariate Behavioral Research*, *46*(2), 340–364. https://doi.org/10.1080/00273171.2011.564527

Velicer, W. F. (1976). Determining the number of components from the matrix of partial correlations. *Psychometrika*, *41*(3), 321–327. https://doi.org/10.1007/BF02293557