

abstract: We consider statistical inference for high-dimensional approximate factor models. We posit a weak factor structure, in which the factor loading matrix can be sparse and the signal eigenvalues may diverge more slowly than the cross-sectional dimension, N . We propose a novel inferential procedure to decide whether each component of the factor loadings is zero or not, and prove that this controls the false discovery rate (FDR) below a pre-assigned level, while the power tends to unity. This "factor selection" procedure is primarily based on a de-sparsified (or debiased) version of the WF-SOFAR estimator of Uematsu and Yamagata (2020), but is also applicable to the principal component (PC) estimator. After the factor selection, the re-sparsified WFSOFAR and sparsified PC estimators are proposed and their consistency is established. Finite sample evidence supports the theoretical results. We apply our procedure to the FRED-MD macroeconomic and financial data, consisting of 128 series from June 1999 to May 2019. The results strongly suggest the existence of sparse factor loadings and exhibit a clear association of each of the extracted factors with a group of macroeconomic variables.