Statistica Sinica **36** (2026), 1-26
doi:https://doi.org/10.5705/ss.202023.0219

SPARSE FACTOR MODEL FOR HIGH DIMENSIONAL TIME SERIES

Xiaoran Wu, Baojun Dou and Rongmao Zhang*

Zhejiang University, City University of Hong Kong and Zhejiang Gongshang University

Abstract: Factor models have been extensively employed in high dimensional time series. However, little is known for the case with the sparse loading matrix. This paper introduces a sparse factor model with an easy-to-implement estimation method, aiming to enhance interpretability and relax the constraints on the dimension p of the time series. In particular, it is shown that under weak conditions, the loading space could be consistently estimated with a convergence rate related to the sparseness of each column in the loading matrix and the eigenvalues used to recover the latent factor and loading matrix. In addition, a randomized sequential test is introduced to determine the number of sparse factors. Simulations and real data analysis on sea surface air pressure and stock portfolios are also provided to illustrate the performance of the proposed method.

Key words and phrases: α-mixing, high dimensional time series, orthogonal projection, sparse factor model.

1. Introduction

Modeling high dimensional time series is of great interest and importance in a wide range of fields including signal process, medical research and financial analysis. For interpretability and simplicity, it is commonly assumed that the high dimensional data is driven by a low dimensional latent factor model. Regarding high dimensional factor models that utilize PCA estimator, a strong-factor framework is usually employed for theoretical consistency, which assumes that the leading eigenvalues driven by common factors are proportional to the dimension p, or that the loading matrix is dense with each element being non-vanishing (Fan, Liao and Mincheva, 2013; Lam, Yao and Bathia, 2011). Otherwise, the consistency of the estimator is compromised. For example, Baik and Silverstein (2006), Paul (2007), and Nadler (2008) show that when the largest eigenvalue has a finite upper bound, the leading sample principal eigenvector is asymptotically orthogonal to the leading population principal eigenvector almost surely.

Through the strong factor benefits from a large dimension p, namely the “blessing of dimensionality” (see the convergence rate in Lam, Yao and Bathia (2011) for example), it suffers less interpretability as the dimension p increases,

*Corresponding author. E-mail: rmzhang@zjgsu.edu.cn