Accepted Manuscript

Asymptotics for empirical eigenvalue processes in high-dimensional linear factor models

Lajos Horváth, Gregory Rice

PII: S0047-259X(17)30436-0
DOI: https://doi.org/10.1016/j.jmva.2018.07.001
Reference: YJMVA 4380

To appear in: *Journal of Multivariate Analysis*

Received date : 23 July 2017

Please cite this article as: L. Horváth, G. Rice, Asymptotics for empirical eigenvalue processes in high-dimensional linear factor models, *Journal of Multivariate Analysis* (2018), https://doi.org/10.1016/j.jmva.2018.07.001

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ACCEPTED MANUSCRIPT

Asymptotics for empirical eigenvalue processes in high-dimensional linear factor models

Lajos Horváth[a], Gregory Rice[b]

[a]*Department of Mathematics, University of Utah, Salt Lake City, UT 84112–0090 USA*
[b]*Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1*

Abstract

When vector-valued observations are of high dimension N relative to the sample size T, it is common to employ a linear factor model in order to estimate the underlying covariance structure or to further understand the relationship between coordinates. Asymptotic analyses of such models often consider the case in which both N and T tend jointly to infinity. Within this framework, we derive weak convergence results for processes of partial sample estimates of the largest eigenvalues of the sample covariance matrix. It is shown that if the effect of the factors is sufficiently strong, then the processes associated with the largest eigenvalues have Gaussian limits under general conditions on the divergence rates of N and T, and the underlying observations. If the common factors are "weak", then N must grow much more slowly in relation to T in order for the largest eigenvalue processes to have a Gaussian limit. We apply these results to develop general tests for structural stability of linear factor models that are based on measuring the fluctuations in the largest eigenvalues throughout the sample, which we investigate further by means of a Monte Carlo simulation study and an application to US treasury yield curve data.

Keywords: Change point analysis, Linear factor models, Principal component analysis

1. Introduction

The largest eigenvalues and corresponding eigenvectors of the sample covariance matrix of vector-valued observations are frequently used as a simplified summary of their variability and dependence structure, especially if the dimension N of the observations is large. This general practice is commonly referred to as principal component analysis (PCA), and within the last two decades there has been a surge of research in both the statistics and econometrics communities aiming to understand the asymptotic properties of PCA when N is large in relation to the sample size T; for a brief summary of this literature we refer to [8, 15, 24, 35]. These asymptotic analyses are often considered within the framework of linear factor models, which basically posit that the dependence between coordinates may be explained by a common linear dependence on a small number of random "factors". Under such models and as $N \to \infty$, the largest eigenvalues of the covariance matrix diverge assuming that the factor loadings do not shrink too much as N increases. This observation seems to have given rise to more general spiked covariance models, and asymptotics for the largest sample eigenvalues under such models are considered in [36, 47]. An important distinction in each of these works is the relative rate at which N may increase with respect to T in order for the asymptotics to hold.

By and large most papers in this direction assume the vector-valued observations form a simple random sample, although in many arenas of application, such as finance and econometrics, the data are observed as time series that are potentially serially correlated or conditionally heteroscedastic. To give a few examples, PCA and the largest eigenvalues are often utilized in Markowitz portfolio optimization [32, 33], and to model co-movements of markets and stocks as a barometer for risk [28, 53], among many other applications. Moreover, with such time ordered data, it is often also of interest to determine before conducting such analyses whether or not the second order structure measured by the largest eigenvalues is homogeneous throughout the sample, or if it appears instead to exhibit one or more structural breaks. If the data under consideration consist of US macroeconomic indicators, for example, then the onset of a recession or the introduction of a new technology may be evidenced by instability in the largest eigenvalues of the covariance matrix. Additionally, PCA based forecasting methods might be improved if changes in the second order structure of the data are taken into account.

Download English Version:

https://daneshyari.com/en/article/11029713

Download Persian Version:

https://daneshyari.com/article/11029713

Daneshyari.com