Discrepancy between global and local principal component analysis on large-panel high-frequency data

Xinbing Kong

Nanjing Audit University
E-mail: xinbingkong@126.com

Abstract: The global principal component analysis (GPCA), PCA applied to the whole sample, is not reliable to reconstruct the common components of a large-panel high-frequency data when the factor space is time-varying, but it works when the factor space does not change in the time domain. The local principal component analysis (LPCA), PCA carried on subsamples, results in consistent estimates of the common components even if the factor loading processes follow continuous-time It^{o} semimartingales, but it loses efficiency when the factor space is time invariant. This motivates us to study the discrepancy between the GPCA and LPCA in recovering the common components of a large-panel high-frequency data. In this paper, we measure the discrepancy by the total sum of squared differences between common components reconstructed from GPCA and LPCA. The asymptotic distribution of the discrepancy measure is provided when the factor space is time invariant and the dimension p and the sample size n tends to infinity simultaneously. Alternatively when some factor loadings are time-varying, the discrepancy measure explodes in a rate higher than $sqrt{pk^{3/2}_n/n}$ under some mild signal conditions on the magnitude of time-variation of the factor loadings, where k_n is the size of each subsample. We apply the theory to test the hypothesis that the factor space does not change in time. We show that the test performs well in controlling the type I error and detecting time-varying factor spaces. This is checked by extensive simulation studies. A real data analysis provides strong evidence that the factor space is always time-varying within a time span longer than one week.