

(Online) Divide-and-Conquer: A Distributed Hierarchical Factor Approach to Modeling Large-Scale Time Series Data

Date: 17 November, 2021 (Wed)

Time: 10:00am - 11:00am
(Hong Kong Time)



SCAN HERE

Online registration:

https://www.cityu.edu.hk/sdsc_web/zoom/

Prof Ruey S. Tsay

H.G.B. Alexander Professor of **Econometrics & Statistics**

Booth School of **Business, University of Chicago**

Abstract

This paper proposes a hierarchical approximate-factor approach to analyzing high-dimensional, large-scale heterogeneous time series data using distributed computing. The new method employs a multiple-fold dimension reduction procedure using Principal Component Analysis (PCA) and shows great promises for modeling large-scale data that cannot be stored nor analyzed by a single machine. Each computer at the basic level performs a PCA to extract common factors among the time series assigned to it and transfers those factors to one and only one node of the second level. Each 2nd-level computer collects the common factors from its subordinates and performs another PCA to select the 2nd-level common factors. This process is repeated until the central server is reached, which collects factors from its direct subordinates and performs a final PCA to select the global common factors. The noise terms of the 2nd-level approximate factor model are the unique common factors of the 1st-level clusters. We focus on the case of 2 levels in our theoretical derivations, but the idea can easily be generalized to any finite number of hierarchies, and the proposed method is also feasible for the data stored and to be analyzed by a single machine with heterogeneous and multilevel subcluster structures. We discuss some clustering methods when the group memberships are unknown and introduce a new diffusion index approach to forecasting. We further extend the analysis to unit-root nonstationary time series. Asymptotic properties of the proposed method are derived for the diverging dimension of the data in each computing unit and the sample size T . We use both simulated and real examples to assess the performance of the proposed method in finite samples, and compare our method with the commonly used ones in the literature concerning the forecasting ability of extracted factors.



Biography

Ruey S. Tsay is H.G.B. Alexander Professor of Econometrics & Statistics, Booth School of Business, University of Chicago. He earned his PhD from the University of Wisconsin - Madison and was with the Carnegie Mellon University before joining University of Chicago in 1989. His research interest includes financial econometrics, analysis of high-dimensional dependent data, forecasting, machine learning, and time-series analysis. He served as co-editor of the Journal of Business and Economic Statistics from 1995 to 1997, Journal of Forecasting from 2006-2013, and Statistica Sinica from 2014-2017. Professor Tsay published widely in leading econometric and statistical journals. He is the author of Analysis of Financial Time Series (3rd ed., 2010, Wiley), An Introduction to Analysis of Financial Data with R (2013, Wiley), and Multivariate Time Series Analysis (2014, Wiley), and co-author of Nonlinear Time Series Analysis (with R. Chen, 2018, Wiley) and Statistical Learning of Big Dependent Data (with D. Peña, 2021, Wiley). He is an elected member of Academia Sinica, Taiwan, and a fellow of the American Statistical Association and the Institute of Mathematical Statistics. He also serves on academic advisory committee of several research institutes.

