

Essays on Statistical and Machine Learning

Methods for Dependent Data Analysis

Ziyan Zhao

Otaru University of Commerce

January 7, 2023

Abstract

This dissertation consists of three independent chapters surrounding statistical and machine learning methods developed to solve some problems in dependent data analysis.

In the first chapter, which is co-authored with Qingfeng Liu¹, we propose a time-varying structural approximate dynamic factor (TVS-ADF) model by extending the ADF model in state-space form. The TVS-ADF model considers time-varying coefficients and a time-varying variance-covariance matrix of its innovation terms, so that it can capture complex dynamic economic characteristics. We also propose an effective Markov chain Monte Carlo (MCMC) algorithm to estimate the TVS-ADF. To avoid the overparameterization caused by the time-varying characteristics of the TVS-ADF, we include the shrinkage and sparsification approaches in the MCMC algorithm. Extensive artificial simulations demonstrate that the TVS-ADF has better forecast performance than the ADF in almost all settings for different numbers of explained variables,

¹I undertake the main work of this research. This study was presented in the Asian meeting of the econometric society in China, 2022.

numbers of explanatory variables, sparsity levels, and sample sizes. An empirical application to macroeconomic forecasting also indicates that our model can substantially improve predictive accuracy and capture the dynamic features of an economic system better than the ADF.

In the second chapter, which is co-authored with Qingfeng Liu and Masamune Iwasawa², we propose a tying maximum likelihood estimation (TMLE) method to improve the performance of estimation of statistical and econometric models in which most time series have long sample periods, whereas the other time series are very short. The main idea of the TMLE is to tie the parameters of the long time series with those of the short time series together by introducing some restrictions on parameters so that some useful information can be transferred from the long series to the short series, which can help improve the estimation accuracy of parameters tied. We first provide asymptotic properties of the TMLE and show its finite-sample risk bound under a fixed tuning parameter which determines the strength of tying. In addition, we provide a bootstrap procedure to select the tuning parameter. Then a finite-sample theory about this bootstrap procedure is developed, which tells us how to conduct the bootstrap procedure effectively. Extensive artificial simulations and empirical applications show that the TMLE has an outstanding performance in point estimate and forecast.

In the third chapter, I provide empirical insight into the impact of the COVID-19 pandemic on the consumer price index (CPI) using a difference-in-difference approach³. Using monthly panel data for eight CPI categories for China and considering two specifications (i.e., the average effect and month-by-month effect), we reveal that the pandemic had a persistent negative impact on housing and daily consumables, whereas no evidence was found for a strong ef-

²All authors contribute equally to this work.

³This study has already been accepted by *The Economic Review*, 73(4).

fect on health care. Regarding education, culture, and recreation, the pandemic mainly had a persistent positive effect over the initial months of the pandemic and then a negative effect for several months. In addition, the pandemic could have a positive effect on food, tobacco, and liquor, while it may have a persistent negative impact on clothing, transport, and communications. Furthermore, there could be a positive effect, which has increased slightly since the pandemic outbreak, on other articles and services.