Waseda Workshop on Time Series Analysis and Related Statistical Methods

Organizer: Yan Liu (Waseda University)
Date: January 13~15, 2024
Venue: Sugadaira seminar house (**菅平セミナーハウス 早稲田**)
Address: 1223-58 Sugadaira-kogen, Ueda, Nagano, 3862200, Japan
https://www.waseda.jp/inst/student/facility/seminar/facility/sugadaira

This workshop is supported by

- JSPS KAKENHI Kiban (C): Grant-in-Aid for Scientific Research No. 23K11018 (Y. Liu)
- Waseda Univ. Tokutei-Kadai: 2023C-094 (Y. Liu)

(* the support covers the accommodation costs.)

Program

January 13

Session I (19:30 – 21:30): chaired by Gaku Miyazawa

19:30-19:50 (1)
Classification of Wildfire Images Based on Machine Learning Classification Models
Dichuan Zheng (Waseda University)

19:50-20:10 (2)
Introduction to Fusion Learning
Motoki Ono (Waseda University)

20:10–20:30 (3)
Practical Approaches to Assessing Portfolio Risk
Wangzhengtian Su (Waseda University)

20:30-20:50 (4)
V-statistic for high-dimensional time series
Yosei Yoshida (Waseda University)

20:50-21:10 (5)
Some discussions on projection pursuit autoregression
Jiahao Guo (Waseda University)

21:10-21:30 (6)
Correlation tests for dependent circular data
Taiga Ohashi (Waseda University)

January 14

Session II (9:30 – 11:45): chaired by Yan Liu

9:30–10:30 (7)
Regression analysis for covariate-adaptive randomization: A robust and efficient inference perspective
Hanzhong Liu (Tsinghua Univ.)

10:45-11:45 (8)
Sparse Spatio-temporal Autoregressions
Yingying Ma (Beihang Univ.)

January 15: 9:30 - 12:00

Integrated Discussion and Research Collaboration

Abstract:

(1)
Classification of Wildfire Images Based on Machine Learning Classification Models
Dichuan Zheng (Waseda University)

Abstract:
In recent years, there has been a globally increasing number of wildfires that cause great damage to ecological stability, environmental health, and economic development. The advances in machine learning, however, allow us to enhance the effectiveness of wildfire detection and classification. We present a comprehensive study of the application of a variety of machine learning models on a wildfire image data set, along with the process of model selection aiming to improve the accuracy and reliability of wildfire classification. We employ a wildfire image dataset obtained from Kaggle, specifically created for training and evaluating machine learning models. The collection has been carefully chosen to encompass a diverse range of photographs portraying wildfires, which have been classified into two distinct kinds to facilitate their categorization and analysis. The 'type0' category comprises photographs that do not depict wildfires, including diverse natural landscapes such as forests, grasslands, and urban environments, all lacking any visual indications of fire. Conversely, 'type1' refers to the dataset that includes wildfires, depicting various views of active fires, smoke, and the environmental devastation caused by wildfires. These photos depict many wildfire scenarios, encompassing diverse meteorological conditions, different times of day (day and night), and various geographical areas. In addition to individual model training and evaluation, we also explore the possibility of model ensembling. By combining the predictions of multiple models, we aim to

leverage their collective strengths, thereby enhancing the overall performance of models on wildfire image classification. Ensemble skills are particularly effective in addressing the weaknesses of individual models and providing more robust predictions. In order to contribute to the development of reliable tools for wildfire classification, the technologies of data augmentation are applied in this task, including affine transformation, multiply alters, and linear contrast adjustments.

(2)
Introduction to Fusion Learning
Motoki Ono (Waseda University)

Abstract:
We propose nonparametric fusion Learning, which is a method for integrating multiple inferences. Unlike existing inferences, this method does not require parametric assumptions. The main tool underlying the proposed framework is the new notion of depth confidence distribution (depth-CD), which is developed by combining data depth and confidence distribution. The numerical results from simulated studies show the approach to be less biased and more efficient than the traditional approaches in nonnormal settings.

(3)
Practical Approaches to Assessing Portfolio Risk
Wangzhengtian Su (Waseda University)

Abstract:
This paper delves into the critical domain of risk management by offering a detailed examination of backtesting methodologies applied to standard risk measures, particularly focusing on Value at Risk (VaR) and Expected Shortfall (ES). The central theme revolves around enhancing

the understanding of risk measures through rigorous backtesting frameworks. Through theoretical elaboration and empirical analysis, this paper demonstrates the superiority of the backtesting method. Finally, the implications of these findings for risk management practice are discussed and directions for future research are suggested.

(4)
V-statistic for high-dimensional time series
Yosei Yoshida (Waseda University)

Abstract:
We consider the problem of testing for homoscedasticity in high-dimensional time series, under the assumption that the sample size n and the dimension p satisfy $\frac{p}{n} \to c \in (0, \infty)$ as $n, p \to \infty$. The homoscedasticity refers to the case where the covariance matrix of the time series is equal to the identity. The test statistic is so-called the V-statistic in the multivariate statistics. The asymptotic null distribution of the V-statistic is shown to be asymptotically normal. The simulation study illustrates the finite sample properties of the V-statistic.

(5)
Some discussions on projection pursuit autoregression
Jiahao Guo (Waseda University)

Abstract:
Projection pursuit regression has established its efficacy in addressing the 'curse of dimensionality' in the realm of nonparametric regressions. However, the adaptation of this approach to time series analysis encounters certain challenges. It is recognized that nonlinear autoregressive models may not always exhibit geometrical ergodicity, barring certain constraints on the autoregressive functions

and innovation distributions. Identifying both necessary and sufficient conditions for geometrical ergodicity in such models poses a significant challenge. Recent decades have seen the proposal of sufficient conditions for geometrical ergodicity in various nonparametric autoregressive models. Building upon these foundations, our research focuses on the application of projection pursuit regression into time series analysis.

(6)
Correlation tests for dependent circular data
Taiga Ohashi (Waseda University)

Abstract:
In these days, we came to employ much more types of data for time series analysis. Circular data is an example which is found to be useful to periodic data analysis such as weekly, monthly or seasonal movements, especially. However, the field is still developing enough to inspect some proposed methods. The definition of correlation on the unit circle is already known as on the real line, but we have some problems when we try circular correlation tests because some theorems may not hold while we regard them as common on the real line. In detail, we have already had the theorem that robust correlation and cross-correlation coefficient coefficients converge into some normal distribution with mean zero in the real line in law, but of course, we cannot use normal distributions on the circle straightly, so we must make a substitute in order to make circular theorems before the experiment. As a result, we need three steps following. First of all, we simulate the central limit theorem (CLT) on the unit circle so as to define the model of the convergence in distribution. We must exclude some distributions which do not satisfy the theorem. In addition, it is helpful for us that the model also has reproductive property same as normal distribution because

we are able to measure the convergence with only mean and variance. It is not a necessary condition for this experiment but worth trying testing. Then, we test the convergence of a correlation coefficient (matrix) with the limited models.

(7)
Regression analysis for covariate-adaptive randomization: A robust and efficient inference perspective
Hanzhong Liu (Tsinghua Univ.)

Abstract:
Linear regression is arguably the most fundamental statistical model; however, the validity of its use in randomized clinical trials, despite being common practice, has never been crystal clear, particularly when stratified or covariate-adaptive randomization is used. In this paper, we investigate several of the most intuitive and commonly used regression models for estimating and inferring the treatment effect in randomized clinical trials. By allowing the regression model to be arbitrarily misspecified, we demonstrate that all these regression-based estimators robustly estimate the treatment effect, albeit with possibly different efficiency. We also propose consistent non-parametric variance estimators and compare their performances to those of the model-based variance estimators that are readily available in standard statistical software. Based on the results and taking into account both theoretical efficiency and practical feasibility, we make recommendations for the effective use of regression under various scenarios. For equal allocation, it suffices to use the regression adjustment for the stratum covariates and additional baseline covariates, if available, with the usual ordinary-least-squares variance estimator. For unequal allocation, regression with treatment-by-covariate interactions should be used, together with our proposed variance estimators. These recommendations apply to simple and stratified randomization, and minimization,

among others. We hope this work helps to clarify and promote the usage of regression in randomized clinical trials.

(8)
Sparse Spatio-temporal Autoregressions
Yingying Ma (Beihang Univ.)

Abstract:
We consider a class of spatio-temporal models with sparse autoregressive coefficient matrices and exogenous variable. We first introduce a profiling approach to extract the exogenous variable from the response, which leads to a profiled model structure. Subsequently, an optimal GMM approach can be applied to such profiled model structure. Next, to overcome the over-determined issue, a novel bagging-based estimator is further developed. Some discussions would be proposed to learn the sparse structure of the autoregressive coefficient matrices. A new BIC-type selection criteria is further developed to conduct variable selection for GMM estimators. The proposed methodology is illustrated with extensive simulation studies and some interesting datasets are analyzed for illustration purpose.