Waseda International Symposium

Topological Data Science, Causality & Time Series Analysis

Date: February 25 – 27, 2021 Venue: Nishi-Waseda Campus, Waseda University Building 55N, Meeting Room 2, 1st floor (Access map: https://www.waseda.jp/fsci/en/access/)

Organizer: Masanobu TANIGUCHI (Research Institute for Science & Engineering, Waseda University)

Supported by:

- JSPS KAKENHI Kiban (S) Grand-in-Aid No. 18H05290 (M. Taniguchi)
- Waseda Research Institute for Science & Engineering Institute for Mathematical Science

Waseda International Symposium

"Topological Data Science, Causality & Time Series Analysis"

Date: February 25–27, 2021

Venue: Nishi-Waseda Campus, Waseda University Building 55N, Meeting Room 2, 1st floor

(Access map: https://www.waseda.jp/fsci/en/access/)

Organizer: Masanobu TANIGUCHI

(Research Institute for Science & Engineering, Waseda University)

This workshop is supported by:

- JSPS KAKENHI Kiban (S) Grand-in-Aid No. 18H05290 (M. Taniguchi)
- > Waseda Research Institute for Science & Engineering, Institute for Mathematical Science

Program

February 25

13:20–13:30: Masanobu Taniguchi (Waseda University) *Opening*

Session I (13:30-15:30): chaired by Ying Chen

13:30–14:10: Yan Liu (Waseda University) <u>Statistical Inference for Persistence Landscapes of the Granger Causality</u>

14:10–14:50: Liang-Ching Lin (National Cheng Kung University) ZOOM <u>Symbolic Interval-Valued Data Analysis for Time Series Based on Auto-</u> <u>Interval-Regressive Models</u>

14:50–15:30: Shih-Feng Huang (National University of Kaohsiung) ZOOM *Multi-Asset Empirical Martingale Price Estimators for Financial Derivatives*

15:30-15:40: Coffee break

Session II (15:40–17:55): chaired by Ngai Hang Chan

15:40–16:25: Mike K P So (HKUST)ZOOMBayesian Network Analysis of Systemic Risk in Financial Markets16:25–17:10: David Preinerstorfer (University Libre Brussels)ZOOMFunctional sequential treatment allocation

17:10–17:55: Christian Francq (ENSAE)ZOOMAdaptiveness of the empirical distribution of residuals in semi-parametricconditional location scale models

February 26

Session III (9:30–10:40): chaired by Yuichi Goto

9:30–10:00: Yujie Xue (Waseda University) <u>Pros and cons on the combination of quantile regression and LASSO under</u> <u>long memory</u>

10:00–10:40: Xiaofei Xu (Waseda University) <u>Regularized Partial Functional Autoregressive Model</u>

10:40–10:50: Coffee Break

Session IV (10:50–12:15): chaired by Yan Liu

10:50–11:30: Yuan Wang (University of South Carolina)ZOOMGroup Analysis of Multi-Trial Brain Signals via Persistent Morse Homology

11:30–12:15: Moo K. Chung (University of Wisconsin-Madison)ZOOMExact topological inference on resting-state brain networks

12:15–13:40: Lunch Time

Session V (13:30–14:55): chaired by Xiaofei Xu

13:30–14:10: Xiaoling Dou (Waseda University) Causality for CHARN models

14:10–14:55: Ying Chen (National University of Singapore)ZOOMProbabilistic Forecasting for Daily Electricity Loads and Quantiles for
Curve-to-Curve Regression

14:55–15:15: Coffee Break

Session VI (15:15–16:45): chaired by Masanobu Taniguchi

15:15–16:00: Ching-Kang Ing (National Tsing Hua University) ZOOM <u>Consistent Order Selection for ARFIMA Processes</u>

16:00–16:45: Elvezio Ronchetti (University of Geneva) ZOOM <u>Saddlepoint approximations for short and long memory time series: A</u> <u>frequency domain approach</u>

16:45–16:55: Coffee Break

Special Art Session VII (16:55-18:35): chaired by Ilia Negri

16:55–17:40: Anna Clara Monti (University of Sannio)ZOOMTango: dance and statistical thinking

17:50–18:35: Yuya Harada (Tokyo University of the Arts) *Pianist* III: Sayo Zenyoji (Tokyo University of the Arts)
The vocal performance of OPERA & Japanese artistic songs

18:45– Friendship Event

February 27

Session VIII (9:30-11:40): chaired by Ching-Kang Ing

9:30–10:10: Yuichi Goto (Waseda Univ.) *Likelihood ratio processes under non-standard settings*

10:10–10:55: Ngai Hang Chan (The Chinese Univ. of Hong Kong)ZOOMStatistical Inference for Glaucoma Detection

10:55–11:40: Integrated Discussion

Abstracts

February 25 (13:20-17:55)

Yan Liu

<u>Title: Statistical Inference for Persistence Landscapes of the</u> <u>Granger Causality</u>

Abstract: We propose a topological approach to statistically analyzing the Granger causality. Granger introduced his celebrated new measure of causality in the sense of prediction errors of multivariate time series 50 years ago. We localize his idea and construct a theory based on locally stationary processes for its alternative version, a natural refinement for stationary processes by Hosoya. To construct the theory, we provide a Gaussian approximation of the suprema of empirical spectral processes. Especially, the local extension of the theory serves for the statistical inference for the Granger causality curve. In addition, we provide a bootstrap procedure for the approximation to construct confidence bands. Finally, we discuss the persistence diagrams and persistence landscapes for the causality curves and numerically construct some examples of locally stationary processes for our simulations studies. (Joint work with Masanobu Taniguchi and Hernando Ombao)

Liang-Ching Lin

<u>Title: Symbolic Interval-Valued Data Analysis for Time Series</u> <u>Based on Auto-Interval-Regressive Models</u>

Abstract: This study considers interval-valued time series data. To characterize such data, we propose an auto-interval-regressive (AIR)

model using the order statistics from normal distributions. Furthermore, to better capture the heteroscedasticity in volatility, we design a heteroscedastic volatility AIR (HVAIR) model. We derive the likelihood functions of the AIR and HVAIR models to obtain the maximum likelihood estimator. Monte Carlo simulations are then conducted to evaluate our methods of estimation and con rm their validity. A real data example from the S&P 500 Index is used to demonstrate our method. (Joint work with Hsiang-Lin Chien and Sangyeol Lee)

Shih-Feng Huang <u>Title: Multi-Asset Empirical Martingale Price Estimators for</u> Financial Derivatives

Abstract: This study proposes an empirical martingale simulation (EMS) and an empirical P-martingale simulation (EPMS) as price estimators for multi-asset financial derivatives. Under mild assumptions on the payoff functions, strong consistency and asymptotic normality of the proposed estimators are established. Several simulation scenarios are conducted to investigate the performance of the proposed price estimators under multivariate geometric Brownian motion, multivariate GARCH models, multivariate jump-diffusion models, and multivariate stochastic volatility models. Numerical results indicate that the multi-asset EMS and EPMS price estimators are capable of improving the efficiency of their Monte Carlo counterparts. In addition, the asymptotic distribution serves as a persuasive approximation to the finite-sample distribution of the EPMS price estimator, which helps to reduce the computation time of finding confidence intervals for the prices of multi-asset derivatives.

Mike K P So

<u>Title: Bayesian Network Analysis of Systemic Risk in Financial</u> <u>**Markets**</u>

Abstract: Analyzing systemic risk in financial markets has been an active research area in financial econometrics, risk management and big data analytics. This paper proposes an approach based on network analysis to study the interrelationship between financial companies. We develop statistical models to understand how the financial network, and thus systemic risk, changes over time. We adopt Bayesian inference methods to estimate the financial network, do network prediction and use listed companies in Hong Kong to illustrate our idea.

David Preinerstorfer <u>Title: Functional sequential treatment allocation</u>

Abstract: We consider a setting, in which a policy maker assigns subjects to treatments, observing each outcome before the next subject arrives. Initially, it is unknown which treatment is best, but the sequential nature of the problem permits learning about the effectiveness of the treatments. While the multi-armed-bandit literature has shed much light on the situation when the policy maker compares the effectiveness of the treatments through their mean, economic decision making often requires targeting purpose specific characteristics of the outcome distribution, such as its inherent degree of inequality, welfare or poverty. In this talk we introduce and study sequential learning algorithms when the distributional characteristic of interest is a general functional of the outcome distribution.

Christian Francq

<u>Title: Adaptiveness of the empirical distribution of residuals in</u> <u>semi-parametric conditional location scale models</u>

Abstract: This paper addresses the problem of deriving the asymptotic distribution of the empirical distribution function \$ \widehat{F}_n\$ of the residuals in a general class of time series models, including conditional mean and conditional heteroscedaticity, whose independent and identically distributed errors have unknown distribution \$F\$. We show that, for a large class of time series models (including the standard ARMA-GARCH), the asymptotic distribution of $\operatorname{F}_n(\operatorname{Cdot})$ is impacted by the estimation but does not depend on the model parameters. It is thus neither asymptotically estimation free, as is the case for purely linear models, nor asymptotically model dependent, as is the case for some nonlinear models. The asymptotic stochastic equicontinuity is also established. We develop consequences of these results for the estimation and tests of the innovations quantiles, which have useful applications in econometrics and finance. The paper also contains a numerical study based on simulated and real data. (Joint work with Jean-Michel Zakoïan)

February 26 (9:30-18:35)

Yujie Xue

<u>Title: Pros and cons on the combination of quantile regression</u> and LASSO under long memory

Abstract: In this talk, we apply LASSO (the least absolute shrinkage and selection operator) to linear quantile regression. Considering that there exists the situation that the norm of different column in the regression matrix may have different order of sequence length n, we introduce a modified LASSO estimator where the tuning parameter λ is not a scalar but a vector. Two situations where the dimension of parameters p is fixed, and where p increases as n increases are discussed. Besides, some simulation studies are examined with a comparison with some other approaches in both of the above situations.

Xiaofei Xu <u>Title: Regularized Partial Functional Autoregressive Model</u>

Abstract: Functional time series and high-dimensional scalar predictors frequently arise in a wide range of modern economic and business applications, which require statistical models that can simultaneously handle the temporal and causal dependence that are prevalent in large sets of mixed-type data. We propose a partially functional autoregressive model (pFAR) to describe the dynamic evolution of the serially correlated functional response on its own lagged values and the causal relation with a large amount of exogenous scalar predictors. Our estimation is conducted by facilitating the sieve method and a two-layer sparsity assumption that

is imposed on groups and elements. In the high-dimensional setting, the sparse structure is completely unknown and it is identified entirely data-driven with a forward-looking criterion. In addition, asymptotic properties of the estimators are established. Extensive simulation studies show that the pFAR model accurately identifies the sparse structure with a convincing and stable predictive performance. The power of the pFAR model is further confirmed by real data analysis of day-ahead gas demand and supply curve predictions of multiple nodes in the German natural gas transmission network with different functions. Given the historical values of the daily curves and 85 scalar predictors, the model detects several essential categories of mixed-type predictors with insightful economic interpretation. It also provides appealing out-of-sample forecast accuracy when compared to a number of popular alternative models. (Joint work with Ying Chen and Thorsten Koch)

Yuan Wang

<u>Title: Group Analysis of Multi-Trial Brain Signals via Persistent</u> <u>Morse Homology</u>

Abstract: Topological data analysis (TDA) can decode multiscale patterns in electroencephalographic (EEG) signals not captured by standard temporal and spectral features. A challenge for applying TDA to groups of long EEG recordings is the ambiguity of performing statistical inference and computational efficiency. To address this problem, we advance a unified inference framework based on a fast permutation test for comparing the TDA descriptor persistence landscape (PL) between two groups of multi-trial EEG signals. The topological inference framework is applied to investigate the EEG correlates of speech sensorimotor impairment in post-stroke aphasia patients under a speech altered auditory feedback (AAF) paradigm. Our analysis reveals a significant difference between the PL features extracted from the event- related potential (ERP) response in aphasia vs. control groups over the parietal-occipital and occipital regions when there is no pitch shift in the auditory feedback and over the parietal region when there an upward pitch shift. The findings validate the application of TDA analysis as a robust tool for investigating the neural correlates of speech sensorimotor impairment in neurological patients suffering from speech-language disorders.

Moo K. Chung <u>Title: Exact topological inference on resting-state brain networks</u>

Abstract: Advances in functional magnetic resonance imaging (fMRI) enable us to measure spontaneous fluctuations of neural signals in the brain in higher spatial and temporal resolution than before. Many previous studies on resting-state fMRI have mainly focused on the topological characterization of graph theory features that fluctuate over the choice of parameters. Persistent homology provides a more coherent mathematical framework for quantifying brain networks that are robust to changes. Instead of looking at networks at a fixed scale, persistent homology charts the changes in topological features such as Betti numbers over every possible parameter. In doing so, it reveals the most persistent topological features that are robust to parameter and noise. In this talk, the exact probability distribution on the Betti numbers that are used in determining the statistical significance will be discussed (Network Neuroscience 3:674-694). Two open mathematical problems (three-sample test and higher order Betti number) related to Betti numbers will be presented.

Xiaoling Dou <u>Title: Causality for CHARN models</u>

Abstract: The CHARN model was proposed in financial data analysis. Because of its non-normality, non-linearity and the blindingly general form, it has come into use in various fields of time series. We consider a nonlinear causality for the CHARN models. We show that the causality of the CHARN models can be evaluated by a Portmanteau test, based on a constrained maximum likelihood estimator of the parameters, and the test statistic has an approximate asymptotic Chisquare distribution. We describe the Chi-square Asymptotics of the Portmanteau test for a CHARN model, provide calculations of the test statistic and investigate the performance of the Portmanteau test by simulation.

Ying Chen

<u>Title: Probabilistic Forecasting for Daily Electricity Loads and</u> <u>Quantiles for Curve-to-Curve Regression</u>

Abstract: Probabilistic forecasting of electricity load curves is of fundamental importance for effective scheduling and decision making in the increasingly volatile and competitive energy markets. We propose a novel approach to construct probabilistic predictors for curves (PPC), which leads to a natural and new definition of quantiles in the context of curve-to-curve linear regression. There are three types of PPC: a predict set, a predictive band and a predictive quantile, and all of them are defined at a pre-specified nominal probability level. In the simulation study, the PPC achieve promising coverage probabilities under a variety of data generating mechanisms. When applying to one day ahead forecasting for the French daily electricity load curves, PPC outperform several state-ofthe-art predictive methods in terms of forecasting accuracy, coverage rate and average length of the predictive bands. For example, PPC achieve up to 2.8-fold of the coverage rate with much smaller average length of the predictive bands. The predictive quantile curves provide insightful information which is highly relevant to hedging risks in electricity supply management. (Joint work with Xiuqin Xu, Yannig Goude and Qiwei Yao. Available at <u>https://arxiv.org/ abs/2009.01595</u>)

Ching-Kang Ing <u>Title: Consistent Order Selection for ARFIMA Processes</u>

Abstract: Estimating the orders of the autoregressive fractionally integrated moving average (ARFIMA) model has been a longstanding challenge in time series analysis. This study tackles the challenge by establishing the consistency of the Bayesian information criterion (BIC) in the ARFIMA model with independent errors. Since we allow the model's memory parameter to be any unknown real number, our consistency result can apply simultaneously to shortmemory, long-memory, and non-stationary time series. We further extend BIC's consistency to the ARFIMA model with conditional heteroskedastic errors, thereby broadening the criterion's range of applications. Finally, the finite-sample implications of our theoretical results is illustrated using numerical examples.

Elvezio Ronchetti

<u>Title: Saddlepoint approximations for short and long memory</u> <u>time series: A frequency domain approach</u>

Abstract: Saddlepoint techniques provide numerically accurate, small sample approximations to the distribution of estimators and test statistics. Except for a few simple models, these approximations are not available in the framework of stationary time series. We contribute to fill this gap. Under short or long range serial dependence, for Gaussian and non Gaussian processes, we show how to derive and implement saddlepoint approximations for two relevant classes of frequency domain statistics: ratio statistics and Whittle's estimator. We compare our new approximations to the ones obtained by the standard asymptotic theory and by two widely-applied bootstrap methods. The numerical exercises for Whittle's estimator show that our approximations yield accuracy's improvements, while preserving analytical tractability. A real data example concludes the paper. (Joint work with Davide La Vecchia)

Anna Clara Monti <u>Title: Tango: dance and statistical thinking</u>

Abstract: The talk briefly recalls the origins of Tango, reviews the main styles and types of music and illustrates the fundamental steps and figures. Tango does not rely on well established sequences of steps but is heavily based on improvisation. At any moment, the dance depends on the connections among the dancers, but it is also influenced by the music, the style, the expertise of the dancers and their previous interactions. The harmony of a tango relies on how these information are processed by the dancers. During a tango event, dancers alternate active and idling times, and typically they change their partner. A model is outlined to describe the activity of a female dancer. Some videos end the talk.

Yuya Harada

Pianist Sayo Zenyoji

Title: The vocal performance of OPERA & Japanese artistic songs

Abstract: (i) Play multiple operas with different age of different performance expressions.

(ii) Listen to the performance and consider whether statistical consideration can be done.

February 27 (9:30-11:40)

Yuichi Goto <u>Title: Likelihood ratio processes under non-standard settings</u>

Abstract: In this talk, we investigate the likelihood ratio (LR) process for curved normal families, simultaneous equation systems, and oneway random ANOVA models. First, we show that curved normal families and simultaneous equation systems, which are naturally endowed with curved structures in the regression matrix, have local asymptotic normality (LAN) property. On the other hand, the LR process of the one-way random ANOVA model converges to atypical asymptotic distributions driven by the contiguity orders when the variance of the random effect belongs on the boundary of parameter space. When the variance of the random effect belongs to the interior of parameter space, the LR process converges to 0. Consequently, one-way random ANOVA models do not hold LAN property. Therefore, by the Neyman-Pearson framework, we show the test based on the log-likelihood ratio is shown to be asymptotically most powerful. (Joint work with Kaneko, Kojima and Taniguchi)

Ngai Hang Chan <u>Title: Statistical Inference for Glaucoma Detection</u>

Abstract: This talk will review some of the statistical techniques used in the analysis and detection of one of the most commonly encountered eye diseases, Glaucoma. By means of some of the recent artificial intelligence algorithms, ophthalmologists are employing modern machine learning technologies such as CNN and its variates to assist in early detection of eye symptoms in Glaucoma studies. This talk will address some of the related issues and discuss the potential of statistical ideas in these studies. Some examples will be given.