**Kouchi International Seminar**

**“Recent Developments of Quantile Method, Causality and High Dim Statistics ”**

**Date: March 3-5, 2018**

**Venue: Tosa Royal Hotel**

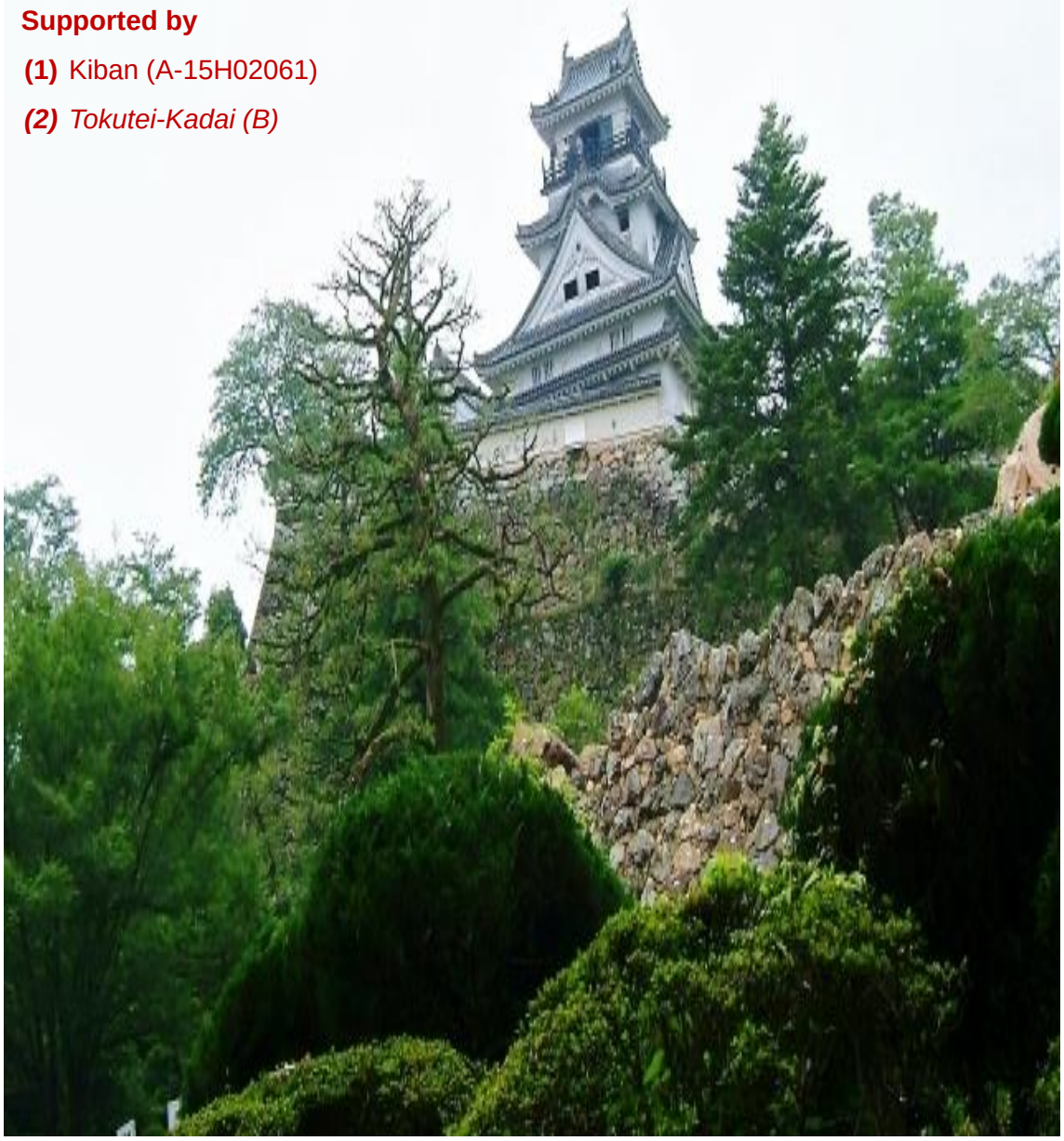
**Organizer:**

Masanobu TANIGUCHI

**Supported by**

**(1)** Kiban (A-15H02061)

***(2)*** *Tokutei-Kadai (B)*



# Kouchi International Seminar

## "Recent Developments of Quantile Method, Causality and High Dim Statistics "

**Date: March 3-5, 2018**

**Venue: Tosa Royal Hotel**
http://www.daiwaresort.jp/tosa/

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

**Supported by**

**(1)** Kiban (A-15H02061) M. Taniguchi, Research Institute for Science & Engineering, Waseda University
***(2)*** *Tokutei-Kadai (B) M. Taniguchi, Research Institute for Science & Engineering, Waseda University*

## Program

**March 3**

20:00-20:30: Applications of Deep Learning in Finance
Ruey S. Tsay, Booth School of Business, University of Chicago

20:30-20:45: Analysts of variance for high dimensional time series
Hideaki Nagahata*(Waseda Univ.). and Masanobu Taniguchi

20:45-21:00: LASSO estimators for high-dimensional time series with long-memory disturbances
Yujie Xue*(Waseda Univ.) and Masanobu Taniguchi

21:15-21:30: Asymptotic theory and numerical studies of Whittle estimation for high-dimensional time series
Yoshiyuki Tanida*(Waseda Univ.), Fumiya Akashi and Masanobu Taniguchi

21:30-21:45: Cox's proportional hazards model with a high-dimensional and sparse regression parameter
Kou Fujimori*(Waseda Univ.)

21:45-22:00: Statiscal inference for weather prediction and weather risk swapping
Makoto Mimizuka* (Waseda Univ.) and Masanobu Taniguchi

**March 4.**

9:45-10:15: Local asymptotic power of self-weighted GEL method and choice of weighting function
Fumiya Akashi*(Waseda Univ.)

10:15-10:45: A nonparametric functional clustering of mouse ultrasonic vocalization data
Xiaoling Dou*(Waseda Univ.)

10:45-11:00: Coffee Break

11:00-11:30: Asymptotic Properties of Mildly Explosive Processes with Locally Stationary Disturbance

Junichi Hirukawa(Niigata Univ.) and Sangyeol Lee

11:30-12:00: Detection of change points in Poisson INAR Models
Hiroshi Shiraishi*( Keio Univ.)

12:00- 13:30: Lunch
13:30 - 14:00:Test of Ambient Fine Particles and Human Influenza in Taiwan: Age group-specific Disparity and Geographic Heterogeneity
Cathy W.S. Chen*(FCU), Ying-Hen Hsieh, Hung-Chieh Su, and Jia Jing Wu

14:00-14:30: From spiked models to factor models: the needle and the haystack
Marc Hallin*( Univ. libre de Bruxelles)

14:30- 15:00: A Dynamic Model of Vaccine Compliance: How Fake News Undermined the Danish HPV Vaccine Program
Peter Hansen*(Univ. North Carolina)

15:00- 15:30: Coffee Break

15:30-16:00: Distribution of baleen whales and predatory fish in relation to available prey in the Norwegian high Arctic
Hiroko Kato Solvang*( *Institute of Marine Research, Bergen)*

16:00- 16:30: COGARCH models: some applications in finance
Ilia Negri*( Univ. Bergamo)

16:30- 17:00: Clustering Data by Extreme Kurtosis Projections
Daniel Peña*(Univ. Carlos III de Madrid), Javier Prieto and Carolina Rendón

**March 5.**

9:30-10:30: Future Developments in Statistics & Research Collaborations
Chaired by Masanobu Taniguchi*(Waseda Univ.)

## Abstracts

March 3 (20:00-22:00)

**Ruey S. Tsay**
**Title: Applications of Deep Learning in Finance**

Abstract: We demonstrate the applications of deep learning in finance via studying the prediction of price changes in high-frequency trading such as transaction-by-transaction intraday trading. Real examples are used in the demonstration.

**Hideaki Nagahata* and Masanobu Taniguchi**
**Title: Analysts of variance for high dimensional time series**

Abstract: For independent observations, analysis of variance (ANOVA) has been enoughly tailored. Recently there has been much demand for ANOVA of high dimensional and dependent observations in many fields. However ANOVA for high dimensional and dependent observations has been immature. In this paper, we study ANOVA for high dimensional and dependent observations. Specifically, we show the asymptotics of classical tests proposed for independent observations and give a sufficient condition

for them to be asymptotically normal. Some numerical examples for simulated and radioactive data are given as applications of these results.

**Yujie Xue* and Masanobu Taniguchi**

**Title: LASSO estimators for high-dimensional time series with long-memory disturbances**

Abstract: LASSO is a $L^1$ norm penalty method to shrink the parameters. Considering the norm of different column with respect of the covariate matrix may have different order of sample size, we introduce modified LASSO estimator where the penalty coefficient λ is not a scalar but vector. Here we discuss the properties of estimator of linear model with long-memory disturbances where the dimension of parameter increases with sample size which is regarded as high dimensional case. It is shown that under some assumption, the sign of LASSO estimators are same with the sign of real parameter with the probability converging to 1 as sample size goes to infinity, and especially when the dimension of parameter has the small order of sample size, the consistency of estimator holds. Joint work with Taniguchi, M..

**Yoshiyuki Tanida*(Waseda Univ.), Fumiya Akashi and Masanobu Taniguchi**

**Title: Asymptotic theory and numerical studies of Whittle estimation for high-dimensional time series**

Abstract: In this presentation, we develop the estimation theory for Whittle functional of high-dimensional non-Gaussian dependent processes. Using a sample version based on a thresholded periodogram matrix, we introduce a thresholded Whittle estimator of unknown parameter, and elucidate its asymptotics. It is shown that the thresholded Whittle estimator is a $\sqrt{n}$ -consistent estimator of the unknown parameter, and that the standardized version has the asymptotic normality. Some numerical studies illuminate an interesting feature of the results. Concretely, for high-dimensional AR(2), we compared the difference of RMSE between the

usual Whittle estimator $\hat{\theta}_w$ and the thresholded estimator $\hat{\theta}_{w,th}$, leading to a conclusion that $\hat{\theta}_{w,th}$ is better than $\hat{\theta}_w$.

**Kou Fujimori**
**Title: Cox's proportional hazards model with a high-dimensional and sparse regression parameter**

Abstract: This talk deals with the proportional hazards model proposed by D. R. Cox in a high-dimensional and sparse setting for a regression parameter. To estimate the regression parameter, the Dantzig selector is applied. The variable selection consistency of the Dantzig selector for the model will be proved. This property enables us to reduce the dimension of the parameter and to construct asymptotically normal estimators for the regression parameter and the cumulative baseline hazard function.

**Makoto Mimizuka* and Masanobu Taniguchi**
**Title: Statiscal inference for weather prediction and weather risk swapping**

Abstract: TBA

## March 4 (9:45-17:00)

**Fumiya Akashi**
**Title: Local asymptotic power of self-weighted GEL method and choice of weighting function**

Abstract: Recently, we often observe the heavy-tailed time series data in variety of fields, and it is unfeasible to apply the classical likelihood ratio-based method to such data directly. To overcome the difficulty, this talk constructs the self-weighted generalized empirical likelihood (SW-GEL) statistic for possibly infinite variance processes, and elucidates the local

asymptotic power of the SW-GEL statistic. The self-weighting method proposed by Ling (2005, JRSS) enables us to control effects brought by the infinite variance of underlying time series models. By the self-weighting method, the proposed statistic converges to the non-central chi-square distribution under the local alternatives. This talk also introduces the selection procedure of tuning parameters in self-weights based on the local asymptotic power.

**Xiaoling Dou**

**Title: A nonparametric functional clustering of mouse ultrasonic vocalization data**

Abstract: Mouse ultrasonic vocalization data are studied in various fields of science. However, methods of automatic data classification and clustering of ultrasonic vocalization data remain to be developed. We define smooth non-harmonic mouse ultrasonic vocalization data as functional data by B-spline basis functions and classify them by shape using the modes of the functional principle component scores. A kernel type estimator is used for defining the modes of the functional data.

**Junichi Hirukawa* and Sangyeol Lee**

**Title: Asymptotic Properties of Mildly Explosive Processes with Locally Stationary Disturbance**

Abstract: In this talk the limit distribution of the least squares estimator for mildly explosive autoregressive models with locally stationary disturbance is established, which is shown to be Cauchy as in the iid case. The result is then applied to identify the onset and the end of an explosive period of a financial time series. Simulations and data analysis are conducted to demonstrate the validity of the result.

Hiroshi Shiraishi

**Title: Detection of change points in Poisson INAR Models**

Abstract: In this study, we consider on-line procedures for detecting changes in the parameters of integer valued autoregressive models of order one. We examine the feasibility of the detector statistics introduced by S. Hudecova et al. (2015,2017). We also propose a criterion to decide a parameter in the test statistics by using ROC (Receiver Operating Characteristic) curves.

**Cathy W.S. Chen*, Ying-Hen Hsieh, Hung-Chieh Su, and Jia Jing Wu**

**Title: Test of Ambient Fine Particles and Human Influenza in Taiwan: Age group-specific Disparity and Geographic Heterogeneity**

Abstract: Influenza is a major global public health problem, with serious outcomes that can result in hospitalization or even death. We investigate the causal relationship between human influenza cases and air pollution, quantified by ambient fine particles less than 2.5μm in aerodynamic diameter ($PM_{2.5}$). A modified Granger causality test is proposed to ascertain age group-specific causal relationship between weekly influenza cases and weekly adjusted accumulative $PM_{2.5}$ from 2009 to 2015 in 11 cities and counties in Taiwan. We examine the causal relationship based on posterior probabilities of the log-linear integer-valued GARCH model with covariates, which enable us to handle characteristics of influenza data such as integer-value, lagged dependence, and over-dispersion. The resulting posterior probabilities show that the adult age group (25-64) and the elderly group in New Taipei in the north and cities in southwestern part of Taiwan are strongly affected by ambient fine particles. Moreover, the elderly group is clearly affected in all study sites. Globalization and economic growth have resulted in increased ambient air pollution (including $PM_{2.5}$) and subsequently substantial public health concerns in the West Pacific region.

Minimizing exposure to air pollutants is particularly important for the elderly and susceptible individuals with respiratory diseases.

**Marc Hallin**

**Title: From spiked models to factor models: the needle and the haystack**

Abstract: a short, nontechnical presentation on statistical inference in high dimension---

**Peter Hansen**

**Title: A Dynamic Model of Vaccine Compliance: How Fake News Undermined the Danish HPV Vaccine Program**

Abstract: Increased vaccine hesitancy present challenges for public health and undermines the effort to eradicate diseases such as measles, rubella, and polio. The decline is partly attributed to misconceptions that are shared on social media, such as the (thoroughly debunked) assertion that vaccines can cause autism. Perhaps, more damaging to vaccine uptake are cases where trusted mainstream media run stories that exaggerate the risks associated with vaccines. It is important to understand the underlying causes of vaccine hesitancy, because these may be prevented, or countered in a timely manner by educational campaigns. In this paper, we develop a dynamic model of vaccine compliance that can help pinpoint events that likely disrupted vaccine compliance. We apply the framework to Danish HPV vaccine data, which experienced a sharp decline in compliance following the broadcast of a controversial TV program.

**Hiroko Kato Solvang**

**Title: Distribution of baleen whales and predatory fish in relation to available prey in the Norwegian high Arctic**

Abstract: Institute of Marine Research in Norway conducts a big project called The Strategic Initiative Arctic (SI-Arctic), which aims to map changes

in the Arctic Ocean as the ice recedes. SI-Arctic has carried a trip for four years (2014-2017) to collect the data using the same methodology as under the ecosystem protocols. They map everything from phytoplankton to whales and birds, and environmental factors. I introduce collected data and analysis for the spatial distribution of the baleen whales, the cod and some of the most relevant prey animals.

**Ilia Negri**

**Title: COGARCH models: some applications in finance**

Abstract: One of the reason that suggest to use COGARCH models to fit financial log-return data is due to the fact that they are able to capture the so called stylized facts observed in real data: uncorrelated log-returns but correlated absolute log-return, time varying volatility, conditional heteroscedasticity, cluster in volatility, heavy tailed and asymmetric unconditional distributions, leverage effects. The aims of this paper is to fit the cogarch models to some real financial data sets, estimate the parameters of the models via the prediction based estimating functions and to look at the performance of these estimates.

**Daniel Peña*, Javier Prieto and Carolina Rendón**

**Title: Clustering Data by Extreme Kurtosis Projections**

Abstract: Peña and Prieto (2001) showed that the extreme kurtosis directions of projected data are optimal for finding clusters when the data has been generated by mixtures of two normal distributions with the same covariance matrix. We generalize this result for any number of mixtures of normal distributions and show that the extreme kurtosis directions of the projected data are linear combinations of the optimal discriminant directions. This is an interesting result because the optimum discriminant direction can only be computed when we know the number of mixtures and the parameters of the distributions. Also, we show that, asymptotically, the extreme kurtosis directions split the distributions or clusters into two sets formed by

components projected together. Thus, we end up with two distributions obtained from merging all the groups. This result suggest a binary decision strategy in order to separate the clusters. In each step we check if the data split into two groups or form a single group by comparing the fitting of a single normal distribution with the fitting of a mixture of two normal distributions. In the second case the process continues while in the first one it stops. The good performance of the algorithm is shown through a simulation study and a comparison with several popular cluster methods.