Waseda International Symposium

"Recent Developments in Time series Analysis:

Quantile Regression, High Dimensional Data &

Causality"

Date: February 26 -28, 2018

Venue: Waseda University, Nishi-Waseda Campus, Building 63, 2nd Floor, Room 5 (Access map: <u>http://www.sci.waseda.ac.jp/eng/access</u>)

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

早稲田大学総合理工学研究所研究重点研究

Supported by

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

Waseda International Symposium

"Recent Developments in Time series Analysis: Quantile

Regression, High Dimensional Data & Causality"

Date: February 26 -28, 2018

Venue: Waseda University, Nishi-Waseda Campus, Building 63, 2nd Floor, Room 5 (Access map: http://www.sci.waseda.ac.jp/eng/access)

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

早稲田大学総合理工学研究所研究重点研究

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 (* Speaker)

February 26 Session (I): 13:30-15:20 chaired by Marc Hallin

13:30-13:40 Opening

Masanobu Taniguchi (Waseda Univ.)

- 13:40-14:25 <u>Three topics in nonparametric statistics</u> Yoichi Nishiyama* (Waseda Univ.)
- 14:25-15:10 Backtesting VaR and expeciles with realized scores

Fabio Bellini, Ilia Negri* (Univ. Bergamo) and Mariya Pyatkova

(15:10-15:20 Coffee break)

Session (II): 15:20-17:35 chaired by Ruey Tsay

15:20-16:05 Nonlinear Causality Test in Multivariate Hysteretic Autoregressive Model

Cathy W. S. Chen^{*} (Feng Chia Univ), Hong Than-Thi, and Mike K.P. So

- 16:05-16:50 <u>Statistical approach to topological data analysis</u> Kenji Fukumizu* (Institute of Statistical Mathematics)
- 16:50-17:35 Parameter Estimation with Out-of-Sample Objective Peter Hansen* (Univ. North Carolina)

February 27

Session (III): 9:30-10:30 chaired by Cathy Chen

9:30-10:15 <u>Self-normalized subsampling method for time series</u> regression models with heavy-tailed long-memory noise Fumiya Akashi* (Waseda University)

10:15-11:00 Quantifying causal relationship in complex ecological systems

Hiroko Kato Solvang* (Institue of Marine Research, Bergen) and Sam Subby

(11:00 - 11:10 Coffee break)

Session (IV): 11:10 - 11:45 chaired by Daniel Peña

11:10-11:55 <u>The mixture transition distribution modeling for</u> higher order circular Markov processes

Takayuki Shiohama and Hiroaki Ogata* (Tokyo Metropolitan Univ.)

11:55-12:40 Möbius transformation and a Cauchy family on the sphere

Shogo Kato* (Institute of Statistical Mathematics, Japan) and Peter McCullagh

(12:40-13:40 Lunch)

Session (V): 13:40 - 15:00 chaired by Davy Paindaveine

13:40-14:25 <u>Jackknife multiplier bootstrap: finite sample</u> approximation to the U-process supremum with applications

Kengo Kato* (Univ. of Tokyo)

14:25-15:10 <u>High-Dimensional Statistical Analysis by</u> Non-Sparse Modeling

Makoto Aoshima* (Univ. of Tsukuba) and Kazuyoshi Yata

(15:10-15:20 Coffee break)

Session (VI): 15:20-17:35 chaired by Masanobu Taniguchi

15:20-16:05 <u>High-Dimensional Time Series Analysis: Some</u> <u>Recent Developments</u>

Ruey S. Tsay* (Univ. of Chicago)

16:05-16:50 <u>Cross-Sectional Effects of Common and</u> <u>Heterogeneous Regressors on Asymptotic Properties of Panel</u> <u>Unit Root Tests</u>

Katsuto Tanaka* (Gakushuin Univ.)

16:50-17:35 <u>Quantile Spectral Analysis for Locally Stationary</u> <u>Time Series</u>

Marc Hallin* (Univ. libre de Bruxelles)

(18:00- Buffet Party)

February 28

Session (VII): 9:30-10:20 chaired by Peter Hansen

9:30-10:15 Efficiency in the high-dimensional one-sample location problem,

Davy Paindaveine* (Univ. libre de Bruxelles)

10:15-11:00 Clustering Time series by dependency

A.M. Alonso and Daniel Peña* (Univ. Carlos III de Madrid)

(11:00-11:15 Coffee break)

Session for different field exchange (VIII): 11:15- 12:00 chaired by Ilia Negri

11:15-12:00 <u>Performance expression and statistical</u> consideration in opera

Yuya Harada* (Tokyo Univ. of the Arts, Baritone Singer)

Abstracts

Three topics in nonparametric statistics

Yoichi Nishiyama* (Waseda Univ.)

Abstract: The talk consists of some small results for three topics in nonparametric statistics. First, we prove the impossibility of weak convergence of kernel density estimators to any non-degenerate law in $L^2(\mathbb{R}^d)$ space. Second, we show asymptotic normality and efficiency of some nonparametric estimators for invariant distribution and density of ergodic diffusion processes based on high-frequency data. Last, we propose a rank-type statistic for nonparametric k-sample and change point problems.

Backtesting VaR and expeciles with realized scores

Fabio Bellini, Ilia Negri* (Univ. Bergamo) and Mariya Pyatkova

Abstract: Several statistical functionals such as quantiles and expectiles arise naturally as the minimizers of the expected value of a scoring function, a property that is called elicitability. The existence of such scoring functions gives a natural way to compare the accuracy of different forecasting models, and to test comparative hypotheses by means of the Diebold-Mariano test. A procedure to backtest VaR and expectiles in an absolute sense, as in the standard Basel I backtesting procedure of VaR, is suggested. To this aim, the asymptotic and finite-sample distributions of realized scores for Normal and uniform i.i.d. samples is studied. The empirical power of our procedure is compared on simulated data, with the alternative procedure based on realized identification functions (i.e. in the case of VaR the number of violations). An higher power in detecting at least misspecification in the mean is fonded. Finally a real data example where both backtesting procedures are applied to AR(1)-Garch(1,1) models fitted to SP500 logreturns for VaR and expectiles forecast is given.

Nonlinear Causality Test in Multivariate Hysteretic Autoregressive Model

Cathy W. S. Chen^{*} (Feng Chia Univ), Hong Than-Thi, and Mike K.P. So

Abstract: This study proposes a multivariate hysteretic

autoregressive model that provides a new way to understand a nonlinear multivariate model in which the regime switch may be delayed when the hysteresis variable lies in a hysteresis zone. We make use of this model to test for Granger causality and for examining which variables have information relevant for improving the predictions of the underlying and unobserved hysteresis dynamics that determines the regimes. We employ an adapted multivariate Student-t distribution in the HAR model to allow for a higher degree of flexibility in the degrees of freedom for each variable. The study creates this adapted multivariate Student-t distribution from modifying the scale mixture of a normal representation of the multivariate Student-t distribution. The employed posterior odds ratios overcome the limitations of the classical approach to multiple testing. Both simulated and real examples help illustrate the proposed methods herein. We use the proposed model to investigate the causal relationship between the quarterly GDP rates of the United States and United Kingdom. We further employ it to check lagged dependence among daily PM2.5 level in three districts of Taipei.

Statistical approach to topological data analysis

Kenji Fukumizu* (Institute of Statistical Mathematics)

Abstract: Topological data analysis (TDA) is a recent methodology for extracting topological and geometrical features from complex geometric data structures. Persistent homology, a new mathematical notion proposed by Edelsbrunner (2002), provides a multiscale descriptor for the topology of data, and has been recently applied to a variety of data analysis. In this talk I will introduce a framework of statistical approach to TDA by combining persistence homology and kernel methods. As an expression of persistent homology, persistence diagrams are widely used to express the lifetimes of generators of homology groups. While they serve as a compact representation of data, it is not straightforward to apply standard data analysis to persistence diagrams, since they consist of a set of points in 2D space expressing the lifetimes. We introduce a method of kernel embedding of the persistence diagrams to obtain their vector representation, which enables one to apply any kernel methods in topological data analysis, and propose a persistence weighted

Gaussian kernel as a suitable kernel for vectorization of persistence diagrams. Some theoretical properties including Lipschitz continuity of the embedding are also discussed. I will also present applications to change point detection and time series analysis in the field of material sciences and biochemistry.

Parameter Estimation with Out-of-Sample Objective

Peter Hansen* (Univ. North Carolina)

Abstracts: We study parameter estimation from the sample X, when the objective is to maximize the expected value of a criterion function, Q, for a distinct sample, Y. This is the situation that arises when a model is estimated for the purpose of describing other data than those used for estimation. The motivated for much estimation has this form, with forecasting problems being a prime example. A natural estimator is the innate estimator that maximizes $Q(X, \theta)$ wrt. θ . While the innate estimator has certain advantages, we show that the asymptotically efficient estimator is defined from a likelihood function in conjunction with Q. The likelihood-based estimator is, however, fragile, as misspecification is harmful in two ways. First, the likelihood-based estimator may be inefficient under misspecification. Second, and more importantly, the likelihood approach requires a parameter transformation that depends on the truth, causing an improper mapping to be used under misspecification. The theoretical results are illustrated with two applications comprising asymmetric loss and multi-step forecasting, respectively.

Self-normalized subsampling method for time series regression models with heavy-tailed long-memory noise

Fumiya Akashi* (Waseda University)

Abstract: Recently, we often observe heavy-tailed and long-range dependent time series data in variety of fields. To model such data, if we use heavy-tailed or long-memory processes, the limit distribution of the fundamental statistic is expressed by involved random variables. This aspect makes it difficult to construct the confidence region for unknown parameters. To make a statistical procedure be robust, this talk makes use of self-normalizing and subsampling approach, and constructs a confidence region for the parameter of the model without any prior estimation for intensity of dependence or tail-behavior of the error process. We also check the finite sample performance of the proposed self-normalized subsampling method, and observe that the proposed method is superior to the classical subsampling method.

Quantifying causal relationship in complex ecological systems Hiroko Kato Solvang* (Institue of Marine Research, Bergen) and Sam Subby

Abstract: For complex ecosystems that involve several interacting species, quantifying causal flow directions is fundamental to understanding system behavior, and in predicting future trajectories of the dynamical system. Existing methodologies for quantifying causality have limited application to complex ecosystems, as the methodologies are usually only restricted to pair-wise considerations.

This paper provides a statistical methodology for causality analysis that integrates Granger's causality analysis based on the log-likelihood function expansion (Partial pair-wise causality), and Akaike's power contribution approach over the whole frequency domain (Total causality). Though the initial idea was proposed by [24], it has hitherto not been applied to complex marine ecosystem dynamics.

We first perform a simulation study to verify the efficacy of the methodology using data generated by several multivariate autoregressive, and bivariate nonlinear time series models. The performance of the methodology is also compared to another power contribution approach that is based on the inverse power spectrum. Finally, based on verifying the performances, we apply the methodology as a new numerical algorithm to real data from the Barents Sea, to derive inter-species relationships that define key food web drivers of the ecosystem. Our results show that the proposed methodology is a useful tool in early stage causality analysis of complex feedback systems.

The mixture transition distribution modeling for higher order circular Markov processes

Takayuki Shiohama and Hiroaki Ogata* (Tokyo Metropolitan Univ.)

Abstract: The stationary higher order Markov process for circular data is considered. We employ the mixture transition distribution (MTD) model to express the transition density of the process. The underlying circular transition distribution is based on the Wehrly and Johnson's bivariate circular models. The structure of the circular autocorrelation function is found to be similar to the autocorrelation function of the AR process on the line. The validity of the model is assessed by applying it to a series of real directional data.

Möbius transformation and a Cauchy family on the sphere

Shogo Kato* (Institute of Statistical Mathematics, Japan) and Peter McCullagh

Abstract: We discuss some properties of a Cauchy family of distributions on the sphere. The family is a spherical extension of the wrapped Cauchy family on the circle. It is seen that the spherical Cauchy family is closed under the Möbius transformation on the sphere and that there is a similar induced transformation on the parameter space. Some properties of a marginal distribution of the spherical Cauchy family are obtained such as certain moments and a closure property associated with the real Möbius group. It is shown that the spherical Cauchy family is transformed into a Cauchy family on the Euclidean space via stereographic projection. Maximum likelihood estimation for the spherical Cauchy is considered; the maximum likelihood estimation for the spherical Cauchy family is equivalent to that for the Cauchy family on the Euclidean space, a closed-form expression for the maximum likelihood estimator is available when the sample size is not greater than three, and the asymptotic normality holds for the maximum likelihood estimator.

Jackknife multiplier bootstrap: finite sample approximation to the U-process supremum with applications

Kengo Kato* (Univ. of Tokyo)

Abstract: In this talk, we consider finite sample approximations to the supremum of a non-degenerate U-process of a general order indexed by a function class. We are primarily interested in situations where the function class as well as the underlying distribution change with the sample size, and the U-process itself is not weakly convergent as a process. Such situations arise in a variety of modern statistical problems. We first consider Gaussian approximations, namely, approximate the U-process supremum by the supremum of a Gaussian process, and derive coupling and Kolmogorov distance bounds. Such Gaussian approximations are, however, not often directly usable in statistical problems since the covariance function of the approximating Gaussian process is unknown. This motivates us to study bootstrap-type approximations to the U-process supremum. We propose a novel jackknife multiplier bootstrap (JMB) tailored to the U-process, and derive coupling and Kolmogorov distance bounds for the proposed JMB method. All these results are non-asymptotic, and established under fairly general conditions on function classes and underlying distributions. Key technical tools in the proofs are new local maximal inequalities for U-processes, which may be useful in other contexts. We also discuss applications of the general approximation results to testing for qualitative features of nonparametric functions based on generalized local U-processes. This is a joint work with Xiaohui Chen (UIUC).

High-Dimensional Statistical Analysis by Non-Sparse Modeling Makoto Aoshima* (Univ. of Tsukuba) and Kazuyoshi Yata

Abstract: High-dimensional data often have a non-sparse and low-rank structure which contains strongly spiked eigenvalues. We call it the "strongly spiked eigenvalue (SSE) model". We note that strongly spiked eigenvalues are accurately detected by using new PCA-type techniques. We also note that, under the SSE model, asymptotic normality is not valid because it is heavily influenced by strongly spiked eigenvalues. In order to give a unified treatment of both the SSE models and non-SSE models, we develop data transformation techniques that transform the SSE models to the non-SSE models. Under the non-SSE models, we can ensure high accuracy for inferences by using asymptotic normality.

High-Dimensional Time Series Analysis: Some Recent Developments

Ruey S. Tsay* (Univ. of Chicago)

High-dimensional time series (HTS) have Abstract: become commonplace in the current data-rich environment. The dynamic or serial dependence of the data distinguishes analysis of HTS from the conventional statistical analysis. Methods developed for independent data may encounter difficulties when they are applied to HTS data. In this talk, we introduce some general concepts of HTS and discuss some recent developments in studying the series. We also discuss dealing strong serial dependence. difficulties for with The developments discussed include testing for serial dependence, Lasso-type regression, and factor models. The difficulties encountered include co-integration. over-differencing, and non-stationarity. Real examples are used in demonstration.

[The talk is based on some joint projects with Rong Chen (Rutgers), Yuefeng Han (U. Chicago), and Shiqing Ling (HKUST).]

Cross-Sectional Effects of Common and Heterogeneous Regressors on Asymptotic Properties of Panel Unit Root Tests Katsuto Tanaka* (Gakushuin Univ.)

Abstract: Dealing with nonstationary panel models with the error term following the AR(1) process, we examine cross-sectional effects of regressors on the asymptotic properties of panel unit root tests. For this purpose we consider various types of common and heterogeneous regressors and compute limiting local powers of tests as T goes to infinity for each N, where T is the time series dimension, whereas N is the cross section dimension. Among the unit root tests considered are those based on the OLSE, GLSE, and other residual-based tests. It is theoretically and graphically shown that the existence of common regressors does not affect the asymptotic properties of these tests, although that of heterogeneous regressors

does affect. We also derive the limiting power envelopes of the most powerful invariant tests, which yields the conclusion that the GLSE-based tests are asymptotically efficient, unlike the time series case. It is also shown that the effect of regressors on the envelopes is shown to be the same as in the tests mentioned above.

Quantile Spectral Analysis for Locally Stationary Time Series

Marc Hallin* (Univ. libre de Bruxelles)

Abstract: Classical spectral methods are subject to two fundamental limitations: they only can account for covariance-related serial dependencies, and they require second-order stationarity. Much attention has been devoted lately to quantile (copula-based) em spectral methods that go beyond traditional covariance-based serial dependence features. At the same time, covariance-based methods relaxing stationarity into much weaker local stationarity conditions have been developed for a variety of time-series models. Here, we are combining those two approaches by proposing copula-based spectral methods for locally stationary processes. We therefore introduce a time-varying version of the copula spectra that have been recently proposed in the literature, along with a suitable local lag-window estimator. We propose a new definition of local strict stationarity that allows us to handle completely general non-linear processes without any moment assumptions, thus accommodating our copula-based concepts and methods. We establish a central limit theorem for the new estimators, and illustrate the power of the proposed methodology by means of a simulation study. Moreover, real-data applications demonstrate that the new approach detects important variations in serial dependence structures both across time and across guantiles. Such variations remain completely undetected, and are actually undetectable, via classical covariance-based spectral methods.

Based on joint work with Stefan Birr (Ruhr-Universität Bochum), Holger Dette (Ruhr-Universität Bochum), Tobias Kley (London School of Economics) and Stanislav Volgushev (University of Toronto), Journal of the Royal Statistical Society Series B 79, 1619– 1643

Efficiency in the high-dimensional one-sample location problem Davy Paindaveine* (Univ. libre de Bruxelles)

Abstract: We consider the one-sample location testing problem in a high-dimensional setup where the dimension p goes to infinity with the sample size n. We consider two types of problems, according to whether the direction of the possible shift is specified or not. Under a sphericity assumption, we investigate the null and non-null asymptotic distributions of two natural tests, namely the (spherical version of) the Hotelling test and the spatial sign test. Our results show in particular that these two tests share the same local asymptotic powers under power-exponential distributions, but not under t distributions. Also, the local alternatives that can be detected by both tests may depend on the underlying distribution, which is non-standard. We further partly read our results in terms of contiguity and local asymptotic normality. Throughout, we conduct Monte Carlo experiments to illustrate the finite-sample relevance of the results.

Clustering Time series by dependency

A.M. Alonso and Daniel Peña* (Univ. Carlos III de Madrid)

Abstract: A new way to find clusters in large vectors of time series is presented. We define a measure of similarity between two series, the generalized cross correlation, that takes into account all their cross correlations until some lag k. The measure is a ratio of the determinants of the correlation matrix of the bivariate vector and the two univariate time series. The matrix of similarities among the series based on this measure is used as input of a hierarchical clustering algorithm. The procedure is automatic and can be applied to large data sets. Clustering using this approach will be useful to build Dynamic Factor Models with Group Structure. The procedure is illustrated with some Monte Carlo experiments and a real data example.

Performance expression and statistical consideration in opera Yuya Harada* (Tokyo Univ. of the Arts, Baritone Singer)

Abstract:

(i) Play multiple operas with different age of different performance

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