BOOK OF ABSTRACTS:
EXTREMES AND TIME SERIES, A WORKSHOP ON THE OCCASION OF RICHARD DAVIS' 70th BIRTHDAY

We will have four sessions on Friday, January 20th

Location: Room C03 School of Social Work (SSW) at 1255 Amsterdam Avenue
Zoom link for the day:
https://columbiauniversity.zoom.us/j/99463101061?pwd=WGxPUE81eEFnZzB3dHgrRC9ad
Ktwdz09
Meeting ID: 994 6310 1061
Passcode: 856738

Breakfast: 8:15-9:00am (10th floor SSW)

Session 1: 9:00-10:30am

Serena Ng (Columbia University)

Time Series Estimation of the Dynamic Effects of Disaster-Type Shocks

This paper provides three results for SVARs under the assumption that the primitive shocks are mutually independent. First, a framework is proposed to accommodate a disaster-type variable with infinite variance into a SVAR. We show that the least squares

estimates of the SVAR are consistent but have non-standard asymptotics. Second, the disaster shock is identified as the component with the largest kurtosis. An estimator that is robust to infinite variance is used to recover the mutually independent components. Third, an independence test on the residuals pre-whitened by the Choleski decomposition is proposed to test the restrictions imposed on a SVAR. The test can be applied whether the data have fat or thin tails, and to over as well as exactly identified models. Three applications are considered. In the first, the independence test is used to shed light on the conflicting evidence regarding the role of uncertainty in economic fluctuations. In the second, disaster shocks are shown to have short term economic impact arising mostly from feedback dynamics. The third uses the framework to study the dynamic effects of economic shocks post-covid. (This is joint work with Richard Davis.)

Beth Andrews (Northwestern University)

Integer-valued GARCH modeling

A generalized autoregressive conditionally heteroskedastic (GARCH) model for uncorrelated, integer-valued time series is introduced. Conditioned on past information, these time series observations have a two-sided Poisson distribution with time-varying variance. Positive and negative observations can have an asymmetric impact on conditional variance. Conditions under which the proposed integer-valued GARCH process is stationary, ergodic, and has finite moments are given. We consider maximum likelihood estimation for model parameters, and give the limiting distribution for these estimators when the true parameter vector is in the interior of its parameter space, and when some GARCH coefficients are on the boundary. We use MLE to fit a GARCH model to observed integer-valued financial time series data. This is work with Xiaofei Hu (Snowflake).

Ruey Tsay (University of Chicago)

A General Matrix-Variate Time Series Model

We introduce a general autoregressive-moving average model for matrix-variate time series, study its identifiability and basic properties, and consider model estimation. We also demonstrate its applications via some real examples.

Coffee break: 10:30-11:00am (10th floor SSW)

Session 2: 11:00-12:30pm.

Holger Rootzén (Chalmers University)

Data quality and prediction skill: rainfall trends in the USA

Is climate change making individual extreme rainfall events in the United States more frequent, more intense, or both? We have developed a new method which make it possible to use high quality NOAA annual maxima series to answer this question. Alternatively, the Peaks over Threshold method could be applied directly to the daily GHCN data -- but these are of lower quality, and then also results could be of lower quality. We further construct new scoring rules for extremes and compare these with existing scoring rules and goodness of fit methods to select the models we use to estimate trends. Our results are that extreme daily rainfall events in the USA are becoming more frequent but that there is little evidence of increasing trends in the distribution of sizes of individual extreme daily rainfall events. The trends are strongest in the northeastern United States where, for many measuring stations, the frequency increase exceeds 150% for each 1°C of average temperature increase. Our aim is to inform infrastructure planning, both for protection against high-impact catastrophes and for local planning of roads and sewers. This is joint work with Helga Olafsdottir and David Bolin.

Johan Segers (Université Catholique de Louvain)

Statistical Inference for Hüsler-Reiss Graphical Models Through Matrix Completions

The severity of multivariate extreme events is driven by the dependence between the largest marginal observations. The Hüsler-Reiss distribution is a versatile model for this extremal dependence, and it is usually parameterized by a variogram matrix. In order to represent conditional independence relations and obtain sparse parameterizations, we introduce the novel Hüsler-Reiss precision matrix. Similarly to the Gaussian case, this matrix appears naturally in density representations of the Hüsler-Reiss Pareto distribution and encodes the extremal graphical structure through its zero pattern. For a given, arbitrary graph we prove the existence and uniqueness of the completion of a

partially specified Hüsler-Reiss variogram matrix so that its precision matrix has zeros on non-edges in the graph. Using suitable estimators for the parameters on the edges, our theory provides the first consistent estimator of graph structured Hüsler-Reiss distributions. If the graph is unknown, our method can be combined with recent structure learning algorithms to jointly infer the graph and the corresponding parameter matrix. Based on our methodology, we propose new tools for statistical inference of sparse Hüsler-Reiss models and illustrate them on large flight delay data in the U.S.

Tailen Hsing (University of Michigan)

Estimating the spectral density of a function-valued spatial process

The spectral density of a stationary process characterizes the second-order properties of the process. In this talk, we focus on the estimation of the spectral density of a spatial process taking values in a function space. As is common in spatial statistics, the process is assumed to be observed at irregularly-spaced spatial locations. We consider a lag-window estimator and present its asymptotic properties, including optimality results assuming that the spectral density belongs to a certain class of functions.

Lunch Break: 12:30-2:00pm (10th floor SSW)

Session 3: 2:00-3:00pm

Simon Tavaré (Columbia University)

A probabilistic model for shallow whole-genome single cell DNA sequencing

We are used to thinking of our genomes as sets of two homologous chromosomes, one ”from Mom” and one “from Dad”; thus a typical cell has copy number two. In cancer, things are very different: cells can have many copies of different regions of the genome. This phenomenon, known as copy number variation, is very important for understanding the evolution of many cancer types. We use a new sequencing technology to estimate

the average copy number in 500,000 basepair regions of cancer genomes. In esophageal cancer, very high copy numbers are possible, but our reconstruction algorithms infer copy numbers far smaller than we believe to be true.

To better understand the technology, we performed a number of control experiments, in which wells containing 1,2,3, or 4 cells are used as input. The DNA in each well is sheared, and a subset of the resulting fragments are sequenced. The output of each experiment is a “pile-up”, a list of the number of times each of roughly 2.5 billion base pairs appears in a sequence read. For a single cell, the coverage should be 0, 1, or 2 at each position; for two cells, 0,1,2,3, or 4 and so on. The data are returned as the proportion of the genome with coverage 0, 1, 2 and so on. Because the sequencing is shallow, most positions are not covered. Experimental error means that read depths greater than twice the number of cells do occur. The problem is to see whether we can call the number of cells in a given experiment, given the coverage data. In this talk I will describe a stochastic model for the coverage process, and a start on how it might resolve the calling problem.

Authors: K. Nowicki-Osuch, B. Wesley, K. Dinh and S. Tavaré, Columbia University

Jay Breidt (NORC at University of Chicago)

On Noncausal Autoregressions: Reversibility, Identifiability and Estimation

We review contributions of Richard Davis and his colleagues to the analysis of linear, non-Gaussian time series models, with emphasis on noncausal autoregressive processes. Such models are not identifiable in the Gaussian case but can be identified from sample path properties in the non-Gaussian case, leading to a variety of feasible estimation and order selection methods. The impact of the work and its more recent extensions are described.

Coffee Break: 3:00-3:30pm (next to C03 SSW)

Session 4: 3:30-5:00pm (Online)

Phyllis Wan (Erasmus University Rotterdam)

Frequently asked questions about Richard Davis

To many students outside of Columbia, Richard Davis has been a celebrated author but a mysterious figure. In this talk, we go through some of the most frequently asked questions about Richard Davis by general PhD students.

William Dunsmuir (University of North South Wales)

What I've learnt from Richard

This talk will reflect on the many things I have learnt from knowing Richard since 1979 when we met as young faculty members in Mathematics at MIT. Two main areas of joint research will be highlighted: deriving accurate asymptotic results for estimation of moving averages with a root close to or on the unit circle and development of theory and methods for regression modelling of time series of counts.

Peter Brockwell (Colorado State University)

Working with Richard.

Reminiscences of meeting, working with and travelling with Richard with a few remarks on continuous-parameter time series.

Reception at 10th floor lounge: 5:00-6:30pm

We will have two sessions on Saturday January 21th

Location: Faculty House
Zoom link for the day:
https://columbiauniversity.zoom.us/j/99463101061?pwd=WGxPUE81eEFnZzB3dHgrRC9adktwdz09
Meeting ID: 994 6310 1061
Passcode: 856738

Breakfast: 10:00-10:30am

Session 1: 10:30-12:30pm

Gennady Samorodnitsky (Cornell University)

Kernel PCA for multivariate extremes

We propose kernel PCA as a method for analyzing the dependence structure of multivariate extremes and demonstrate that it can be a powerful tool for clustering and dimension reduction.

Sidney Resnick (Cornell University)

The effect of reciprocity on the heavy tail nature of preferential attachment

Modeling social networks using preferential attachment (PA) presents obvious challenges. Models can be difficult to analyze and only occasionally do simulations from

calibrated models leave a comfortable impression that simulation matches the reality of data. A typical glaring discrepancy is often shown by “reciprocity”, meaning the percentage of directed edges that link to nodes (network users) in both directions. (“You like me and I like you.” “You reference my paper and I reference yours.”) Real data often exhibits higher empirical reciprocity compared to what is given by simulations from traditional preferential attachment.

We discuss how standard preferential attachment often alarmingly induces small reciprocity. Adding a reciprocity feature to the model fixes this but dramatically changes the asymptotic heavy tail limit measure for frequency of nodes with specified (in, out)-degree. Without reciprocity, the limit measure is spread out in the first quadrant. With reciprocity, the standardized limit measure concentrates on a ray from the origin and there is even hidden regular variation.

(The theoretical work on reciprocity uses embedding techniques in Markov branching models and was led by Tiandong Wang, Fudan University, Shanghai.)

Thomas Mikosch (University of Coppenhagen)

About ratios of homogeneous functionals acting on a heavy-tailed time series

We consider a real-valued regularly varying stationary sequence with in- finite variance and construct homogeneous functionals from a sample of this sequence. These include the maximum of the sample, its sum, lp norms, and their powers. First, we prove the joint convergence of these quantities via transform arguments (hybrid characteristic functions, Laplace transforms). This implies the convergence of their ratios, and we also provide characterizations of their limit ratio and calculation of their moments. (This is joint work with O. Wintenberger and N. Matsui)

Claudia Klüppelberg (Technical University Munich)

An Extremal Tour with Richard (Mostly) Through Munich.

Lunch Break: 12:30-2:00pm

Session 2: 2:00-3:30pm

Dan Cooley (Colorado State University)

Transformed-linear Regularly Varying Time Series Models and Extreme Event Attribution

In order to capture the dependence in the upper tail of a time series, we develop non-negative regularly-varying time series models that are constructed similarly to classical non-extreme ARMA models. We define the tail pairwise dependence function (TPDF) to quantify the extremal dependence between two elements of the regularly-varying time series, and use the TPDF to define the concept of weak tail stationarity for regularly-varying time series. To develop our non-negative regularly-varying ARMA-like time series models, we use transformed-linear operations. We show existence and stationarity of these models and develop their properties, such as the model TPDF's. Transformed-linear models fit to hourly windspeed data outperform other models in terms of estimating high quantiles of sums and run length. We use a transformed-linear model to quantify the increased risk of fire danger in Colorado and estimate that risk of a season similar to that seen in 2020 has increased four times over the past half-century.

John Nolan (American University)

Riesz capacity via hitting distribution for stable processes

A method is described for computing alpha-capacity of sets in d-dimensions through the use of simulated continuous time process with isometric stable increments. A method we call Walk-In-Out-Balls is described to simulate points from the equilibrium measure of a set. These points are then used to estimate the Riesz capacity of the set.

Philippe Naveau (LSCE Institut Pierre-Simon Laplace)

Records Analysis in Climate Attribution problems

Numerical climate models are complex and combine a large number of physical processes. They are key tools in quantifying the relative contribution of potential anthropogenic causes (e.g., the current increase in greenhouse gases) on high-impact atmospheric variables like heavy rainfall or temperatures. These so-called climate extreme event attribution problems are particularly challenging in a multivariate context. In addition, global climate models like any in sillico numerical experiments are affected by different types of bias.

In this talk, I will discuss how to combine two statistical theories (counterfactual theory and extreme value theory) to assess causality in the context of extreme event attribution. In addition, the question of uncertainty quantification that remains a challenge in any climate attribution analysis will be explored from various directions. In particular, a simple model bias correction step for records will be described in detail. To illustrate our approach, we infer emergence times in precipitation from the CMIP5 and CMIP6 archives. (This is joint work with Anna Kiriliouk, Paula Gonzalez, Soulivanh Thao and Julien Worms).