Joint PhD Seminar in Statistics, Actuarial and Financial Mathematics

Bremen University,
Université catholique de Louvain, Louvain-la-Neuve,
Carl von Ossietzky University, Oldenburg

May 25 and 26, 2018
Oldenburg, Germany

Welcome

Dear attendees,

welcome to Oldenburg University for the third joint PhD Seminar in Statistics, Actuarial and Financial Mathematics among the universities of Bremen, Louvain-la-Neuve, and Oldenburg. The scope of this seminar is to bring together PhD students in the mentioned fields of study and to foster discussions among you both on topics of your research and on general PhD related questions. Moreover, the opportunity to present your results in an extended format of 35 minutes allows the colleagues from other places to get a deeper insight into current research of the other groups and will hopefully give you valuable feedback for your work on top of your advisors'. Of course, as any pressing deadline, such a presentation will also push forward your work and get you closer to successfully finishing your thesis.

Speakers

Jörg Thomas Best, UO
Charlie Hillner, UB
Nathalie Lucas, UCL
André Lüschen, UB
André Neumann, UB
Florian Pechon, UCL
Marius Pluhar, OL
Jonathan von Schroeder, UB
Antoine Soetewey, UCL
Natalia Sirotko-Sibirskaya, UB
Tino Werner, UO
Janine Witte, UB

Organizing Team

Werner Brannath
Marcus Christiansen
Michel Denuit
Thorsten Dickhaus
Vanessa Didelez
Catherine Legrand
Angelika May
Peter Ruckdeschel

How to reach Oldenburg University

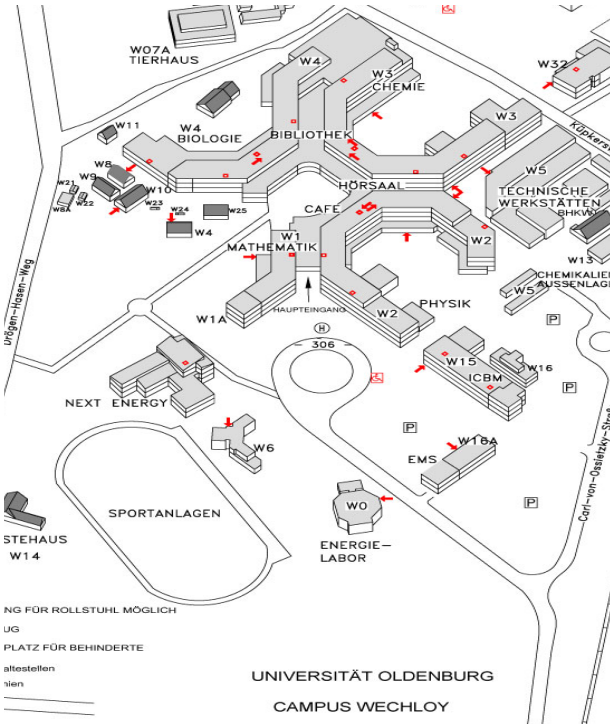
By Car:

Coming from Bremen, take the A28 to Oldenburg and use exit 12, Oldenburg Haarentor, turn right into the Ammerländer Heerstraße and follow the signs to Campus Wechloy, i.e., on Ammerländer Heerstraße, you drive straight on until the Mercedes merchant Rosier to your right, where you turn right into the Carl-von-Ossietzky-Straße. Keep on this street until you reach the parking lot of Campus Wechloy.

GPS: Carl-von-Ossietzky-Straße 9–11, 26129 Oldenburg

By rail and bus:

Coming from Bremen, almost all trains from Nordwestbahn to Oldenburg also call in Oldenburg Wechloy from where it is a 5min walk to the venue of the PhD Seminar. From Oldenburg Hauptbahnhof (main station), you may also take bus 306 which has Wechloy as final destination.



Programme Overview

Contents and Timeline: Friday, May 25, 2018 (Room W03 1-152)

08:50 - 09:00 Welcome — *Marcus Christiansen*, UO

Block 1 — Chair: *Michel Denuit*, UCL

09:00 - 09:35 *Nathalie Lucas*, UCL:
Actuarial models for health and disability insurance

09:45 - 10:20 *Marius Pluhar*, UO:
Finding Markovian models for insurance processes by expanding state spaces

10:30 - 11:00 Coffee Break

Block 2 — Chair: *Werner Brannath*, UB

11:00 - 11:35 *Florian Pechon*, UCL:
MTPL Insurance from a household perspective

11:45 - 12:20 *Charlie Hillner*, UB:
Group sequential designs controlling the population-wise error rate

12:30 - 14:00 Lunch Break (University Cafeteria)

Block 3 — Chair: *Catherine Legrand*, UCL

14:00 - 14:35 *Antoine Soetewey*, UCL:
Life and Health Actuarial Pricing: a Biostatistics Approach

14:45 - 15:20 *André Lüschen* , UB:
Quantification of treatment effect variability and its use for subgroup analyses

15:30 - 16:00 Coffee Break

Block 4 — Chair: *Peter Ruckdeschel*, UO

16:00 - 16:35 *Tino Werner*, UO:
Asymptotic linearity, robustness and sparsity: From regularized regression to ranking

16:45 - 17:20 *Jonathan von Schroeder*, UB:
Efficient Computation of the CDF of Order Statistics

18:30 - 20:30 Dinner at Casa Vecchia,
Kleine Kirchenstraße 8, 23122 Oldenburg

Contents and Timeline: Saturday, May 26, 2018 (Room W03 1-152)

Block 5 — Chair: *Thorsten Dickhaus*, UB

09:00 - 09:35 *Janine Witte*, UB:
Generalised IDA - estimating causal effects from cohort data under model uncertainty

09:45 - 10:05 *Natalia Sirotko-Sibirskaya*, UB:
Frequency-Domain Model Selection in Dynamic Factor Models

10:15 - 10:45 Coffee Break

Block 6 — Chair: *Angelika May*, UO

10:45 - 11:05 *André Neumann*, UB:
Multivariate multiple test procedures based on Bernstein copula estimation

11:15 - 11:50 *Jörg Thomas Best*, UO:
Examination of the L^2-closedness of a space of stochastic integrals

12:00 - 12:10 Closing Session — *Marcus Christiansen*, UO

12:30 - 13:30 Lunch Break at AliBaba,
Ammerländer Heerstraße 120, 26129 Oldenburg

Abstracts

Examination of the L^2-closedness of a space of stochastic integrals

Jörg Thomas Best

Oldenburg University

11:15 - 11:50, Saturday, May 26

This paper deals with the following question arising in Mathematical Finance: "What is the good notion of the space of attainable claims in $L^p(P)$?"
We fix $1 \leq p \leq \infty$ and a semi-martingale $X = (X_t)_{t\in\mathbb{R}}$ which is locally in $L^p(P)$. Furthermore we define the space $G_T(\Theta_p^s)$ of simple p-attainable claims as random variables $\int_0^T HdX$ for a simple predictable integrand H that satisfies some integrability conditions ensuring that we have $\int_0^T HdX \in L^p(P)$. After defining the closure K_p of $G_T(\Theta_p^s)$ in $L^p(P)$ we will study the duality relation between the space K_p and the equivalent martingale measures $\mathcal{Q}$ for X with $\frac{d\mathcal{Q}}{d\mathcal{P}} \in L^q(\mathcal{Q})$, where $\frac{1}{p}+\frac{1}{q}=1$. In addition we will also establish a duality relation between K_p and $\mathcal{M}^{q,s}$, all signed martingale measures $\mathcal{Q}$ with $\frac{d\mathcal{Q}}{d\mathcal{P}} \in L^q(P)$, where $p \geq 1$, $q \geq 1$ and $\frac{1}{p}+\frac{1}{q}=1$.

Group sequential designs controlling the population-wise error rate

Charlie Hillner

Bremen University

11:45 - 12:20, Friday, May 25

In clinical trials one is often concerned with testing hypotheses in several subpopulations of an overall population, usually defined by biomarkers. While these subpopulations are disjoint in many cases, e.g. if the overall population is divided into a biomarker positive and negative group, they can also intersect. If one hypothesis is to be tested in each of these intersecting populations, multiplicity adjustments have to be made. But since the commonly used family-wise error rate (FWER) turns out to be overly conservative in this case, another more liberal error measure is proposed – the population-wise error rate (PWER). To illustrate how to make inference using the PWER we primarily consider the case of two intersecting populations P_1 and P_2, where the mean efficacy of a treatment is tested in each one of them by means of respective hypotheses H1 and H2. Here, an important question is how to deal with $P_1 \cap P_2$ if, for example, H1 is rejected but H2 not. For this purpose, it is shown how group sequential designs

that control the PWER and allow to enrich the intersection $P_1 \cap P_2$ after the interim analysis are constructed. These ideas are then generalized to the K-stage case and extended by an error-spending approach. Since the group sequential approach does not allow design changes throughout the trial, it is discussed in which situations an adaptive design is a more suitable choice.

Actuarial models for health and disability insurance

Nathalie Lucas

Université catholique de Louvain

09:00 - 09:35, Friday, May 25

The main purpose of the research is to develop i) efficient actuarial models of pricing and reserving and ii) risk classification techniques for health insurance products. Whereas classical approach treats health insurance using life techniques, we aim to better study health claims randomness and integrate longevity and other systematic risks (i.e. inflation) in the analysis. The special dependence between morbidity and mortality will be emphasized. The last section will introduce dynamic heterogeneity in health claims using a Hidden Markov Model (HMM).

Quantification of treatment effect variability and its use for subgroup analyses

André Lüschen

Bremen University

14:45 - 15:20, Friday, May 25

The evaluation of the efficacy and safety of a new treatment via a randomized controlled trial is typically accompanied by subgroup analyses to investigate whether the treatment effect in important subgroups (for example defined by sex or age) is consistent with the treatment effect in the whole study population. Subgroup analyses can therefore give stronger evidence for the positive effect of the new treatment and are consequently an integral part of the approving process. However, the topic of subgroup analyses is intensively discussed, see for example Brookes et al. (2001), Pocock et al. (2002) or Wang et al. (2007). A central issue is that subgroup analyses usually have low power to detect inconsistencies when present, which leads to difficulties in the interpretation of these analyses.
We want to contribute to the discussion by proposing a novel approach to detect such heterogeneity. The method is based on the concept of treatment effect variance (TEV). A high TEV may point to substantial uncertainty and to an incon-

sistent treatment effect. We suggest interpreting the observed treatment effect (overall or in subgroups) in the light of the TEV, since the former is of less practical value when the TEV is high. We will show simulation results to illustrate the potential utility of the proposed method.

Multivariate multiple test procedures based on Bernstein copula estimation

André Neumann

Bremen University

10:45 - 11:05, Saturday, May 26

Multivariate multiple test procedures have received growing attention recently. This is due to the fact that data generated by modern applications typically are high-dimensional, but possess pronounced dependencies due to the technical mechanisms involved in the experiments. Hence, it is possible and often necessary to exploit these dependencies in order to achieve reasonable power. We express dependency structures in the most general manner, namely, by means of copula functions. In this talk, we extend previous statistical results regarding bivariate Bernstein copulas to the multivariate case and study their impact on multiple tests. This extends a similar approach by Stange et al. (2015) in the parametric case.

MTPL Insurance from a household perspective

Florian Pechon

Université catholique de Louvain

11:00 - 11:35, Friday, May 25

In Non-Life Insurance, actuarial risk classification are generally confined to univariate, policy-based analyses: individual claim frequencies are modelled for every single product, without accounting for interactions between coverages bought by members of the same household. Models use information about the policyholder, his or her vehicle and the characteristics of the contract in order to predict the claim frequency. However, some residual effects remain unobserved and will cause overdispersion. By proper inclusion of random effects in Poisson model for claim frequencies, this unexplained heterogeneity can be accounted for and will allow for periodic revaluations based on previous claim experience.
In most cases, the insurer has also at his disposal data available to match policyholders from the same household. We aim to use this information to develop a multivariate Poisson mixture model that combines the claim frequencies of all

policyholders from the same household. The unexplained heterogeneity for policyholders from the same household become dependent, which allows periodic revaluation of the claim frequencies estimates based on the whole household's previous claim experience. By using information related to other policyholders from the same household, one increases the complexity of the model but also the amount of information available for each policyholder. This allows refining the prediction of claim frequencies. This model can be used for commercial and strategic decisions by the insurer and possible cross-selling opportunities can be identified.

Finding Markovian models for insurance processes by expanding state spaces

Marius Pluhar

Oldenburg University

09:45 - 10:20, Friday, May 25

In insurance practice, stochastic processes are often assumed to be Markovian. This assumption allows easier calculations and provides a broad and well-known theory. While the dataset used to calibrate the model often does not satisfy the Markov assumption, one can ensure Markovianity by expanding the state space. To evaluate how to expand the state space most efficently and when to stop expanding, a measure for Markovianity is needed. I will present a first idea on how to test a nonstationary, discrete sampled process for Markovianity.

Efficient Computation of the CDF of Order Statistics

Jonathan von Schroeder

Bremen University

16:45 - 17:20, Friday, May 25

In multiple testing many important methods (e.g. FWER control using Holm's step-down procedure and FDR control using the Benjamini-Hochberg (BH) step-up procedure) can be naturally formulated in terms of the order statistics of the p-values. To understand the properties (e.g. the statistical power) of these methods the joint distribution of order statistics is thus of great interest. In the i.i.d. case it can be computed exactly using Steck's recursion, which is however computationally intensive. Thus in the literature either asymptotic approximations are used or only a (very) small number of hypotheses is considered. Motivated by this the talk details how principles from dynamic programming and a careful implementation in the R language can greatly improve the performance of a concrete

implementation. Furthermore, it is shown how these techniques can be applied to a generalisation of Steck's recursion with the goal of enabling an analysis of the finite-sample performance of the BH procedure for a number of hypotheses much larger than previously reported in the literature.

Life and Health Actuarial Pricing: a Biostatistics Approach

Antoine Soetewey

Université catholique de Louvain

14:00 - 14:35, Friday, May 25

It is generally thought that patients having suffered from a cancer have a lower probability of survival compared to healthy people. Due to this aggravated risk and the relatively small number of patients wishing to take out insurance coverage in case of death, the insurance industry is reluctant to grant such a guarantee. However, survival and life expectancy of cancer patients have been increasing over the last decades and we can reasonably assume that it will keep increasing in the future thanks to medical and technological progress. In regard to this, France passed a law referred as "the right to forget", that is, the right for a person subscribing to a contract not to declare a previous cancer after a period of 10 years after the end of the therapeutic protocol (Sapin and Touraine, 2017). This period being reduced to 5 years if the person is a minor. But some questions remain: The thresholds of 10 and 5 years are arbitrary and does not reflect survival of the persons having suffered from a cancer. There remains some ambiguity about what is considered as treatment, so what marks the end of a therapeutic protocol and in the end when the patient will start to benefit from this right? Finally, this right is very binary and not flexible at all.
The aim of the project is twofold: (i) To develop a method to adequately estimate the threshold after which cancer patients can be considered as cured, and (ii) to find a proper way to adapt the actuarial pricing of life insurance products to each category of risk, disease, person, etc. The goal is also to demonstrate that for some types of cancer, the survivors actually have a chance of survival comparable to that of the general population, or pose a moderately increased risk and could therefore be covered in the event of death. This involves measuring and quantifying the potential excess mortality so that the premiums claimed reflect the risk in terms of financial services.

Frequency-Domain Model Selection in Dynamic Factor Models

Natalia Sirotko-Sibirskaya

Bremen University

09:45 - 10:05, Saturday, May 26

We propose a data-driven method for selecting an "optimal" number of factors in the dynamic factor model context. The method is based on Wold-type cross-validation technique in the frequency domain, Wold (1978). In the spirit of Hurvich and Zeger (1990) we define a frequency-domain-cross-validation criterion (FDCV) for a dynamic factor model. It can be shown that the expectation of a FDCV is approximately equal to the mean squared error (MSE) of an estimate of the common part and the variance of idosyncratic components. The criterion is evaluated subsequently for each possible choice of the number of common factors. We study the theoretical properties of the suggested method as well as its performance both in Monte-Carlo simulations and empirically.

Asymptotic linearity, robustness and sparsity: From regularized regression to ranking

Tino Werner

Oldenburg University

16:00 - 16:35, Friday, May 25

The ranking problem intends to predict the right ordering of the response variables in a data set. Practical relevance includes applications in fraud detection, medicine or finance. Our main goal is to construct a parametric ranking method that both provides robustness against wrong model specifications and sparseness in the sense of selecting only the most relevant predictor variables. However, robustifying the loss function for the ranking would result in computational intractability of the optimization problem.
In general, an estimation procedure which results from the solution of a minimization problem can be identified with a suitable M-functional. Such M-functionals are highly non-linear. But in fact, if the functional satisfies some regularity properties which in our case will be compact differentiability, then an infinite-dimensional Delta-method provides an asymptotic linear expansion in terms of influence curves up to some error term depending on the number of observations in the data set. For a suitably consistent starting estimator, this linearization replaces solving optimization problems by evaluating these influence curves at the given data points.
We show under which conditions the asymptotic linear expansion is valid. Furthermore, we provide concrete examples of machine learning algorithms that fit into this framework. The adaptation to an efficient algorithm for the ranking problem will be cast as an open problem.

Generalised IDA - estimating causal effects from cohort data under model uncertainty

Janine Witte

Bremen University

09:00 - 09:35, Saturday, May 26

If a causal effect is to be estimated from observational data, confounding needs to be adjusted for. The adjustment set is best determined based on subject-matter knowledge, e.g. using a known directed acyclic graph (DAG) representing the measured variables. For the case that such knowledge is not available, Maathuis et al. (2009) developed a method called IDA. IDA consists of two steps: In the first step, the causal structure underlying the data is inferred using causal search. The result is a set of DAGs compatible with the data. In the second step, for each of these DAGs a valid adjustment set is determined and a causal effect is estimated by multiple regression, resulting in a multiset of estimates. IDA assumes (among other things) that the measured variables follow a multivariate normal distribution. This might be plausible for genetics data, and indeed IDA has been successfully applied in this field. However, IDA in its original form cannot be applied to other types of data, say data from an epidemiological cohort study where some variables are categorical and some non-normal continuous and the relationships between them are not in general linear. In the presentation, I will sketch a modified version, generalised IDA, that may overcome some of the restrictions of original IDA, and illustrate this using real data.