

Statistische Woche 2019

Book of Abstracts

10. - 13. September 2019
Universität Trier



© Sheila Werner



 **Universität Trier**


Trier
statistische woche **2019**
10. - 13. September

BOOK OF ABSTRACTS

STATISTISCHE WOCHE

10.-13. September 2019

Trier

Deloitte.

DISTATIS
Statistisches Bundesamt

microm
Consumer Marketing


Bitte ein Bit
Bitburger

www.statistische-woche.de

Organisers / Organisatoren:

University of Trier
Universität Trier



German Statistical Society (DStatG)
Deutsche Statistische Gesellschaft (DStatG)



German Municipal Statisticians (VDSt)
Verband Deutscher Städtestatistiker (VDSt)



German Association for Demography (DGD)
Deutsche Gesellschaft für Demographie (DGD)



Italian Statistical Society (SIS)
Società Italiana di Statistica (SIS)



Address / Adresse:

Universität Trier
Universitätsring 15
54296 Trier

STATISTISCHE WOCHE

10.-13. September 2019

Trier – Germany

www.statistische-woche.de

Contents

Programme – Overview	p. 3
Programme	p. 5
• Session Timetable	p. 5
• Tuesday	p. 9
• Wednesday	p. 17
• Thursday	p. 25
• Friday	p. 32
Abstracts	p. 35
Participants	p. 183
Building Plans	p. 192

Sponsors – Sponsoren:

- Deloitte
- DESTATIS – Statistisches Bundesamt
- Lidl
- microm – Consumer Marketing
- Pearson
- RatSWD – Rat für Sozial- und Wirtschaftsdaten
- Springer Verlag
- Statistik der Bundesagentur für Arbeit
- Bitburger

Inhaltsverzeichnis

Programm – Übersicht	S. 3
Programm	S. 5
• Sitzungszeitplan	S. 5
• Dienstag	S. 9
• Mittwoch	S. 17
• Donnerstag	S. 25
• Freitag	S. 32
Abstracts	S. 35
Teilnehmer	S. 183
Gebäudepläne	S. 192

Programme Committee / Programmkomitee:

German Statistical Society / Deutsche Statistische Gesellschaft

Wolfgang Schmid (Chair / Vorsitz DStatG, Europa-Universität Viadrina)

Roland Fried (TU Dortmund)

Ralf Münnich (Universität Trier)

Yarema Okhrin (Universität Augsburg)

Ulrike Rockmann (Senatsverwaltung für Inneres und Sport Berlin)

Peter Schmidt (Statistisches Bundesamt)

Markus Zwick (Statistisches Bundesamt)

Association of German Urban Statisticians / Verband Deutscher Städtestatistiker:

Michael Haußmann (Chair / Vorsitz, Landeshauptstadt Stuttgart, Statistisches Amt)

Hermann Breuer (Conference commissioner / Tagungsbeauftragter VDSt, Stadt Köln, Amt für Stadtentwicklung und Statistik)

German Society for Demography / Deutsche Gesellschaft für Demographie:

Gabriele Doblhammer (Chair / Vorsitz, Universität Rostock)

Martin Kohls (Bundesamt für Migration und Flüchtlinge)

Michaela Kreyenfeld (Hertie School of Governance)

Olga Pötzsch (Statistisches Bundesamt)

Norbert F. Schneider (BIB)

Annelene Wengler (Robert Koch Institut)

Christina Westphal (Fraunhofer-Institut)

Organiser / Organisator:

Ralf Münnich (Universität Trier)

Programme Overview / Programmübersicht:

Tuesday, September 10, 2019:

08:00-09:00	Welcome Coffee	▷ Foyer Building C
09:00-10:40	Sessions	
10:40-11:10	Coffee Break	
11:10-12:00	Opening	▷ Audimax
12:00-12:50	Plenary Session <i>Microsimulations</i> (Gijs Dekkers)	▷ Audimax
12:50-14:10	Lunch Break	
14:10-15:50	Sessions	
15:50-16:20	Coffee Break	
16:20-18:00	Sessions	
18:30-20:00	Poster Presentations	▷ Foyer Building C
19:00-20:00	Board Meeting DStatG	▷ C402

Wednesday, September 11, 2019:

09:00-10:40	Sessions	
10:40-11:10	Coffee Break	
11:10-12:00	Plenary Session <i>Environmental Statistics</i> (Peter Filzmoser)	▷ Audimax
12:00-12:50	Heinz-Grohmann Lecture (Bernd Fitzenberger)	▷ Audimax
12:50-14:10	Lunch Break	
14:10-15:50	Sessions	
15:50-16:20	Coffee Break	
16:20-18:00	Sessions	
18:00-19:00	General Assembly Meeting DStatG	▷HS5

Thursday, September 12, 2019:

09:00-10:40	Sessions	
10:40-11:10	Coffee Break	
11:10-12:00	Plenary Session <i>Housing</i> (Marietta Haffner)	▷ Audimax
12:00-12:50	Gumbel Lecture (Axel Bücher)	▷ Audimax
12:50-14:10	Lunch Break	
14:10-15:50	Sessions	
15:50-16:20	Coffee Break	
16:00-16:20	Springer Best-Poster-Award	▷ Springer booth
16:20-18:00	Sessions	
19:30-22:30	Conference Dinner	▷ Europahalle Trier

Friday, September 13, 2019:

09:00-10:40	R-Tutorial: Small Area Estimation with R	▷ C360
09:00-10:40	Sessions	
10:40-11:10	Coffee Break	
11:10-12:50	Sessions	
11:10-12:00	Plenary Session <i>Data Journalism</i> (Marcel Pauly)	▷ Audimax
12:00-12:30	Panel Discussion <i>Data Journalism</i>	▷ Audimax

Session Timetable / Sitzungszeitplan:

	09:00-10:40	14:10-15:50	16:20-18:00	09:00-10:40	14:10-15:50	16:20-18:00	09:00-10:40	14:10-15:50	16:20-18:00	09:00-10:40	11:10-12:50
Section Session number ▷ Letter of lecture hall	Tue 10/9			Wed 11/9			Thu 12/9			Fri 13/9	
Data Journalism									HS7	HS7	
Environmental Statistics		HS5		HS5							
Microsimulations – Applications and Methods	HS7	HS7	HS7	HS7							
Housing					HS7	HS7					
Bayesian Analysis in Natural and Social Science									C10	C10	
Joint DStatG SIS – Survey Statistics								HS6	HS6		
Young-Academics Mini-Symposium: “Models for Spatial and Spatiotemporal Data”			C10								
Education and Training		C9	C9								
Computational Statistics and Data Analysis		HS5	HS5		HS5	HS5					
Empirical Economics	HS4	HS4	HS4	HS4							
Methodology of Statistical Surveys	HS5					HS6	HS6			HS6	
Regional Statistics	HS6	HS6	HS6	HS6	HS6			HS7			
Statistical Literacy							HS7				
Statistics in Finance							HS4	HS4	HS4	HS4	
Macroeconomics & Finance		C10									
Statistics in Natural Science and Technology								HS5	HS5		
Statistical Theory and Methods					HS4	HS4	HS5				
Nonparametric and Robust Statistics				C10							
Business and Market Statistics				C9			C10	C10			
Labour Market and Social Security						C10					
National Accounts, Welfare Measurement					C10						
DGD: Cause of Death Statistics: Methodological challenges and empirical findings							C9	C9			
FDZ: Recent methodological developments in the FDZ of the statistical offices of the federation and countries					C9						
MAKSWELL EU										HS5	HS5
Verband Deutscher Städtestatistiker		HS6	HS6			C9					

Plenary Sessions & Lectures

Room: Audimax

Tuesday, September 10th, 2019

Plenary Session Microsimulations: **Gijs Dekkers**; Karel Van Den Bosch. „Prospective microsimulation of pensions in European Member States“.

Tuesday, 12:00-12:50 (page 35).

Wednesday, September 11th, 2019

Plenary Session Environmental Statistics: **Peter Filzmoser**. „Correlation analysis for compositional environmental data.“

Wednesday, 11:10-12:00 (page 35).

Heinz-Grohmann Lecture: **Bernd Fitzenberger**. „Lohnungleichheit in Deutschland: Fakten, Daten, Analysen“.

Wednesday, 12:00-12:50 (page 36).

Thursday, September 12th, 2019

Plenary Session Housing: **Marietta Haffner**. „Housing market developments and housing affordability in the European Union.“

Thursday, 11:10-12:00 (page 36).

Gumbel Lecture: **Axel Bücher**. „On the block maxima method in extreme value statistics“.

Thursday, 12:00-12:50 (page 37).

Friday, September 13th, 2019

Plenary Session Data Journalism: **Marcel Pauly**. „What data journalism is and how cooperation with scientists and statisticians can work“.

Friday, 11:10-12:00 (page 37).

Panel Discussion Data Journalism.

Friday, 12:00-12:30 (page 38).

Invited Speakers

Tuesday, September 10th, 2019

Michael König. „A Structural Model for the Coevolution of Networks and Behavior“.

Tuesday, 9:00-9:50 (page 42).

Emilio Porcu. „Random Fields evolving temporally over spheres“.

Tuesday, 16:20-17:10 (page 67).

Wednesday, September 11th, 2019

Przemyslaw Biecek. „Explanation, Exploration and Debugging of Predictive Machine Learning Models“.

Wednesday, 14:10-15:00 (page 109).

Dirk Czarnitzki; Paul Hünermund; Nima Moshgbar. „Public Procurement of Innovation: Evidence from a German Legislative Reform“.

Wednesday, 9:00-9:50 (page 98).

Nazarii Salish. „Functional Time Series Analysis“.

Wednesday, 14:10-15:00 (page 111).

Thursday, September 12th, 2019

Jozef Barunik; Michael Ellington. „Asset Pricing using Time-Frequency Dependent Network Centrality“.

Thursday, 9:00-9:50 (page 141).

Alessandra Petrucci; Emilia Rocco. „A proposal to assess the representativeness of non-probability surveys“.

Thursday, 14:10-15:00 (page 143).

Richard Nickl; Jakob Söhl. „Bernstein-von Mises Theorems for Compound Poisson Processes“.

Thursday, 16:20-17:10 (page 162).

Monica Pratesi; Luciana Quattrociocchi; Gaia Bertarelli; Alessandro Gemignani; Caterina Giusti. „Spatial distribution of multidimensional educational poverty in Italy using small area estimation“.

Thursday, 16:20-17:10 (page 164).

Rainer von Sachs; Marion Rebecca; Govaerts Bernadette; Lederer Johannes. „Selecting variables via clustering within regularized prediction in high dimensions“.

Thursday, 14:10-15:00 (page 157).

Friday, September 13th, 2019

Olha Bodnar. „Non-Informative Bayesian Inference for Heterogeneity in a Generalized Marginal Random Effects Meta-Analysis“.

Friday, 9:25-10:15 (page 175).

Tuesday, September 10th, 2019, 08:00 - 09:00

Welcome CoffeeFoyer Building C

Tuesday, September 10th, 2019, 09:00 - 10:40

Microsimulations – Applications and Methods (Session 1)

Chair: Markus Zwick, Room: HS7

Marc Hannappel; Ardian Canolli. „The International Journal of Microsimulation. Eine bibliometrische Netzwerkanalyse über Akteure und Anwendungsfelder“.

Tuesday, 9:00-9:25 (page 39).

Joachim Merz. „Mikrosimulation und die Hochrechnung von Mikrodaten – Erfahrungen und Anwendungen eines informationstheoretischen Ansatzes und dem Programmpaket ADJUST“.

Tuesday, 9:25-9:50 (page 39).

Hanna Brenzel; Markus Zwick. „Mikroanalyse und Georeferenzierung in der amtlichen Statistik“.

Tuesday, 9:50-10:15 (page 40).

Julian Reinhold; Philip Höcker; Sven Brocker; **Rainer Schnell**. „Die Anwendung von Mikrosimulationen zur Bevölkerungsfortschreibung“.

Tuesday, 10:15-10:40 (page 41).

Empirical Economics (Session 1)

Chair: Robert Jung, Room: HS4

Michael König. „A Structural Model for the Coevolution of Networks and Behavior“.

Tuesday, 9:00-9:50 (page 42).

Ines Lindner; Huang Jia-Ping; Bernd Heidegott. „Naïve Learning in Social Networks with Random Communication“.

Tuesday, 9:50-10:15 (page 42).

Fernando Linardi; Cees Diks; **Marco van der Leij**; Iuri Lazier. „Dynamic interbank network analysis using latent space models“.

Tuesday, 10:15-10:40 (page 43).

Methodology of Statistical Surveys: Small Area

Chair: Thomas Zimmermann, Room: HS5

Joscha Krause; Jan Pablo Burgard; Dennis Kreber. „Regularisierte Small Area Modelle zur robusten Schätzung in der Gegenwart unbekannter Messfehler in den Kovariaten“.

Tuesday, 9:00-9:25 (page 44).

Sandra Hadam; Nora Würz; Ann-Kristin Kreuzmann; Timo Schmid. „City Data from LFS and Big Data“ – Schätzung von Indikatoren der Arbeitskräfteerhebung für funktionale städtische Gebiete mit Mobilfunkdaten“.

Tuesday, 9:25-9:50 (page 45).

Tobias Schoch. „Robuste Schätzer für das Fay-Herriot-Modell“.

Tuesday, 9:50-10:15 (page 46).

Qonnita Silvia Makrufa; Ayesha Tantriana. „Analysis of Poverty in a Village Level of Indonesia with Small Area Estimation: Case in Bangkalan District“.

Tuesday, 10:15-10:40 (page 47).

Regional Statistics: Städte im Vergleich

Chair: Lothar Eichhorn, Room: HS6

Ulrike Rockmann. „Bildungsmonitoring Berlin-Mitte – 2te Projektphase“.

Tuesday, 9:00-9:25 (page 48).

Uwe Neumann; Lisa Taruttis. „Sorting in an urban housing market – is there a response to demographic change?“

Tuesday, 9:25-9:50 (page 48).

Teresa Lauerbach. „Die kleinräumige Perspektive: Disparitäten innerhalb von Städten“.

Tuesday, 9:50-10:15 (page 49).

Ramona Voshage; Martin Gornig. „Eine neue Symbiose von Stadt und Industrie? Vergleich regionaler Beschäftigungs- und Gründungsdynamiken in Deutschland“.

Tuesday, 10:15-10:40 (page 50).

Coffee Break: 10:40 - 11:10

Tuesday, September 10th, 2019, 11:10 - 12:50

Opening

Room: Audimax

Opening.

Tuesday, 11:10-12:00.

Plenary Session: Microsimulations

Chair: Ralf Münnich, Room: Audimax

Plenary Session Microsimulations: **Gijs Dekkers**; Karel Van Den Bosch. „Prospective microsimulation of pensions in European Member States“.*Tuesday, 12:00-12:50 (page 35).*

Lunch Break: 12:50 - 14:10

Tuesday, September 10th, 2019, 14:10 - 15:50

Microsimulations – Applications and Methods (Session 2)

Chair: Petra Stein, Room: HS7

Petra Stein; **Dawid Bekalarczyk**; Eva Depenbrock. „Implementing empirical results of panel models with lagged dependent variables and random intercepts into dynamic microsimulation“.*Tuesday, 14:10-14:35 (page 51).***Sabine Zinn**. „Modelling the Unobserved: An Extension of Traditional Microsimulation“.*Tuesday, 14:35-15:00 (page 51).***Joscha Krause**; Jan Pablo Burgard; Ralf Münnich; Simon Schmaus. „Estimation of Transition Probabilities for Spatial Dynamic Microsimulations in the Presence of Regional Heterogeneity“.*Tuesday, 15:00-15:25 (page 52).***Kristina M. Neufang**; Ralf Münnich. „Synthesizing regression models to estimate micro-level data with multiple sources“.*Tuesday, 15:25-15:50 (page 53).*

Computational Statistics, Data Analysis & Environmetrics

Chair: Bernd Bischl, Room: HS5

Nadja Klein; **Thomas Kneib**; Torsten Hothorn. „Multivariate Conditional Transformation Models“.
Tuesday, 14:10-14:35 (page 54).

Hasan Hüseyin Oruc. „An Efficient, Unsupervised Learning Algorithm for Discovering Meaningful Clusters in Data“.
Tuesday, 14:35-15:00 (page 54).

Miryam Sarah Merk. „Adaptive LASSO estimation of the spatial weights matrix“.
Tuesday, 15:00-15:25 (page 55).

Jona Lilienthal; Axel Bücher; Roland Fried. „Regional flood frequency analysis of possibly heterogeneous groups“.
Tuesday, 15:25-15:50 (page 55).

Empirical Economics (Session 2)

Chair: Robert Jung, Room: HS4

Konstantin Görden; Melanie Schienle. „Evaluating Effects of Tuition Fees: Lasso for the Case of Germany“.
Tuesday, 14:10-14:35 (page 56).

Livia Shkoza; Derya Uysal; Winfried Pohlmeier. „Heterogeneity in Network Peer Effects“.
Tuesday, 14:35-15:00 (page 56).

Marc Schneble; Göran Kauermann. „Estimation of Latent Network Flow in Bike-Sharing Systems from Station Feeds“.
Tuesday, 15:00-15:25 (page 57).

Michael Lebacher; Göran Kauermann. „Regression-based Network Reconstruction with Nodal and Dyadic Covariates and Random Effects“.
Tuesday, 15:25-15:50 (page 57).

Regional Statistics: Forum der Regionalstatistik – Neue digitale Daten in der Regional- und Städtestatistik (Session 1)

– Gemeinsame Session von DStatG-Regionalstatistik und VDSt –

Chair: Hartmut Bömermann, Lothar Eichhorn, Room: HS6

Tobias Gramlich. „Neue Daten für die amtliche Statistik: Verknüpfen von Daten aus dem Internet mit Daten einer Erhebung bei Beherbergungsbetrieben in Hessen: Herausforderungen, Methoden und Ergebnisse“.
Tuesday, 14:10-14:35 (page 58).

Tobias Gramlich. „Daten aus Mobilfunknetzwerken für die amtliche Statistik: ein Pilotprojekt in der hessischen Tourismusstatistik“.
Tuesday, 14:35-15:00 (page 59).

Normen Peters. „Identifikation von Unternehmen mit E-Commerce-Aktivitäten in der amtlichen Statistik durch Webscraping und Machine Learning“.
Tuesday, 15:00-15:25 (page 60).

Clara Maria Schartner. „Fernerkundungsdaten in der amtlichen Statistik“.
Tuesday, 15:25-15:50 (page 61).

Education and Training (Session 1)

Chair: Ulrike Rockmann, Room: C9

Florian Berens; Sebastian Hobert. „Kontinuierliches Lernen in Einführungen in die Statistik durch intelligente Systeme unterstützen“.
Tuesday, 14:10-14:35 (page 62).

Florian Berens. „Messung statistischer Kompetenz am Ende eines Einführungskurses in die Statistik“.
Tuesday, 14:35-15:00 (page 62).

Macroeconomics & Finance

Chair: Florian Ertz, Room: C10

Christian Bauer; Marc Adolph. „Structured Eurobonds: Pricing the political cost of joint liability“.
Tuesday, 14:10-14:35 (page 64).

Ulf von Kalckreuth; Norman Wilson. „Towards integrating Fintech into statistical classification systems – a process oriented approach“.
Tuesday, 14:35-15:00 (page 64).

Ulf von Kalckreuth. „Statistical Governance and FDI in Emerging Economies“.
Tuesday, 15:00-15:25 (page 65).

Coffee Break: 15:50 - 16:20

Tuesday, September 10th, 2019, 16:20 - 18:00

Microsimulations – Applications and Methods (Session 3)

Chair: Jan Pablo Burgard, Room: HS7

Tyll Krüger. „Social contagion processes on networks: which random graph model fits best?“
Tuesday, 16:20-17:10 (page 66).

Federico Bonofiglio; Martin Schumacher; Harald Binder. „Distributed Survival Analysis under Disclosure Constraints“.
Tuesday, 17:10-17:35 (page 66).

Young-Academics Mini-Symposium: “Models for Spatial and Spatiotemporal Data”
Chair: Philipp Otto, Room: C10

Emilio Porcu. „Random Fields evolving temporally over spheres“.
Tuesday, 16:20-17:10 (page 67).

Anna Gloria Billé; Francisco Blasques; Leopoldo Catania. „Dynamic Spatial Autoregressive Models with Time-varying Spatial Weighting Matrices“.
Tuesday, 17:10-17:35 (page 67).

Gong Chen; Ostap Okhrin. „Interpolation of Weather Conditions for a Flight Corridor“.
Tuesday, 17:35-18:00 (page 68).

Empirical Economics (Session 3)
Chair: Karsten Schweikart, Room: HS4

Yves Schüler. „How should we filter economic time series?“
Tuesday, 16:20-16:45 (page 69).

Karsten Schweikart. „Oracle Efficient Estimation of Structural Breaks in Cointegrating Regressions“.
Tuesday, 16:45-17:10 (page 69).

Regional Statistics: Forum der Regionalstatistik – Neue digitale Daten in der Regional- und Städtestatistik (Session 2)
– *Gemeinsame Session von DStatG-Regionalstatistik und VDSt* –
Chair: Hartmut Bömermann, Lothar Eichhorn, Room: HS6

Julian Schulz; Franziska Große. „Quo vadis, Statistik? – Eine GIS-gestützte Analyse der niedersächsischen Beherbergungswirtschaft?“
Tuesday, 16:20-16:45 (page 70).

Günther Bachmann. „Neue Pendleranalysen mit Mobilfunkdaten“.
Tuesday, 16:45-17:10.

Education and Training (Session 2)

Chair: Ulrike Rockmann, Room: C9

Kevin Stabenow; Friederike Schmal. „Seminar “Statistical Literacy” – Statistikwissen nutzbar machen“.

Tuesday, 16:20-16:45 (page 72).

Toni Stocker. „Lernzentrierte Übungssystematik und statistisches Monitoring“.

Tuesday, 16:45-17:10 (page 73).

Christine Buchholz. „No-Screen-Policy in der Präsenzlehre – verbunden mit eLearning-Angeboten zum selbstgesteuerten Lernen“.

Tuesday, 17:10-17:35 (page 73).

Karsten Lübke; Matthias Gehrke. „Learning to understand and to do Statistics step-wise by R shiny, learnr, mosaic and markdown“.

Tuesday, 17:35-18:00 (page 74).

Computational Statistics and Data Analysis: Applied Statistics

Chair: Peter Filzmoser, Room: HS5

Christoph Herrmann; Thomas Kirschstein; David Lukas. „Drivers of individual performance of professional biathletes“.

Tuesday, 16:20-16:45 (page 75).

Andreas Behr; Gerald Fugger. „PISA Performance of Natives and Immigrants: Selection versus Efficiency“.

Tuesday, 16:45-17:10 (page 75).

Marius Ötting; Roland Langrock; Christian Deutscher; Vianey Leos Barajas. „Is there a hot hand effect in professional darts? An analysis using state-space models“.

Tuesday, 17:10-17:35 (page 76).

Tuesday, September 10th, 2019, 18:30 - 20:00

Poster Presentations

Room: Foyer Building C

Philipp Breidenbach; Lea Eilers; Jan Fries. „Rent Control and Rental Prices: High Expectations, High Effectiveness?“

(page 77).

Patricia Dörr; Ralf Münnich. „A Machine Learning-based Multivariate Non-Parametric Density Estimator“.

(page 77).

Saeed Fayyaz; Arash Fazeli. „Application of location-based Data Analysis for Census Planning (Case study in Iran)“.
(page 78).

Christoph Frohn; Petra Stein. „Dynamic Microsimulation Modelling of Health among older Migrants in Germany“.
(page 78).

Laura Güdemann; Ralf Münnich. „Sensitivity of Composite Indicators due to Non-Sampling Errors“.
(page 79).

Marc Hannappel; Philipp Jakobs. „Digitale Aktionsraumforschung. Neue Methoden zur Analyse von Bewegungsprofilen im städtischen Raum“.
(page 80).

Tobias Maier; Stefanie Steeg; Gerd Zika. „Der Einfluss berufsstruktureller Informationen auf Stellenbesetzungsprozesse – heute und in der Zukunft?“
(page 81).

Caroline Neuber-Pohl. „Identifizierung von Ausbildungsenden ohne Abschluss in der IEBS“.
(page 82).

Monika Obersneider; Petra Stein. „Developing a Spatial Microsimulation Model for Labor Market Integration of Migrants in Germany“.
(page 83).

Peter Pütz; Stephan B. Bruns. „The (Non-)Significance of Reporting Errors in Economics: Evidence from Three Top Journals“.
(page 83).

Other events on Tuesday

19:00 : Board Meeting DStatG ▷ C402

19:30 : Get-together at the Bitburger Wirtshaus

Wednesday, September 11th, 2019, 09:00 - 10:40

Microsimulations – Applications and Methods (Session 4)

Chair: Hanna Brenzel, Room: HS7

Kerstin Bruckmeier; Regina T. Riphahn; Jürgen Wiemers. „Benefit underreporting in survey data and its consequences for measuring non-take-up: new evidence from linked administrative and survey data“.
Wednesday, 9:00-9:25 (page 84).

Kerstin Bruckmeier; **Jannek Mühlhan**; Jürgen Wiemers. „Adjusting the welfare system to new labour market risks: Integrating an in-work benefit into the German basic income system“.
Wednesday, 9:25-9:50 (page 84).

Sebastian Dräger. „REGIONAL MICROSIMULATION MODEL OF THE FUTURE NEED FOR PRIMARY SCHOOL INFRASTRUCTURE IN TRIER“.
Wednesday, 9:50-10:15 (page 85).

Jana Emmenegger; Ralf Münnich. „Building a longitudinal income module for microsimulations of spatial inequality“.
Wednesday, 10:15-10:40 (page 86).

Environmental Statistics

Chair: Roland Langrock, Room: HS5

Walter J. Radermacher. „30 years environmental metrics – What have we learnt? Where to go?“
Wednesday, 9:00-9:25 (page 87).

Andreas Hense. „Climate Change Detection and Attribution: An applied problem in Environmental Statistics“.
Wednesday, 9:25-9:50 (page 87).

Katharina Hees. „Return times of extreme weather events“.
Wednesday, 9:50-10:15 (page 88).

Marco Oesting; Philippe Naveau. „Spatial Modeling of Heavy Precipitation in the South of France Using Max-Stable Coupling Between Weather Station Recordings and Ensemble Forecasts“.
Wednesday, 10:15-10:40 (page 89).

Empirical Economics (Session 4)

Chair: Gabriel Okasa, Room: HS4

Jamol Bahromov; Vasyl Golosnoy; Markus Pape. „The Rotated Empirical Similarity Approach: Theory and Applications“.
Wednesday, 9:00-9:25 (page 90).

Britta Stöver. „Estimating the transition time from school to university using a stochastic mortality model“.

Wednesday, 9:25-9:50 (page 90).

Philipp Böing; **Bettina Peters.** „Effectiveness and Efficacy of R&D Subsidies: Estimating Treatment Effects with One-sided Noncompliance“.

Wednesday, 9:50-10:15 (page 91).

Michael Lechner; **Gabriel Okasa.** „Random Forest Estimation of the Ordered Choice Model“.

Wednesday, 10:15-10:40 (page 92).

Regional Statistics: Regionale Disparitäten und Strukturen

Chair: Michael Fürnrohr, Room: HS6

Miriam Orlowski. „Einkommensungleichheit zwischen den bayerischen Kreisen 1991 bis 2016“.

Wednesday, 9:00-9:25 (page 93).

Frederik Parton. „Eine Faktoren- und Clusteranalyse der Raumordnungsregionen Deutschlands auf Basis von Projektionsdaten des Modells QMORE für den Zeitraum 2015-2035“.

Wednesday, 9:25-9:50 (page 93).

Arne Lehmann. „Regionale Disparitäten in Niedersachsen – dargestellt anhand einer Clusteranalyse auf Gemeindeebene“.

Wednesday, 9:50-10:15 (page 94).

Antonia Milbert. „Ausprägung und Veränderung regionaler Disparitäten – eine Analyse mittels Markov-Ketten“.

Wednesday, 10:25-10:40 (page 95).

Nonparametric and Robust Statistics

Chair: Christine Müller, Room: C10

Florian Dumpert. „Quantitative Robustheit von Prädiktoren auf Basis lokaler SVMs“.

Wednesday, 9:00-9:25 (page 96).

Dennis Malcherczyk; Kevin Leckey; Christine H. Müller. „K-sign depth: Asymptotic distribution and efficient computation“.

Wednesday, 9:25-9:50 (page 96).

Mirko Alexander Jakubzik. „Semi-parametric and robust estimation in intensity-based models“.

Wednesday, 9:50-10:15 (page 97).

Business and Market Statistics: Research, Education, Innovation

Chair: Andreas Kladroba, Room: C9

Dirk Czarnitzki; Paul Hünermund; Nima Moshgbar. „Public Procurement of Innovation: Evidence from a German Legislative Reform“.*Wednesday, 9:00-9:50 (page 98).***Verena Eckl.** „Auftragsforschung und Innovationsfähigkeit der Unternehmen in Deutschland“.*Wednesday, 9:50-10:15 (page 98).***Rainer Frietsch.** „Patentstatistik – Möglichkeiten, Verlockungen und Fehlinterpretationen“.*Wednesday, 10:15-10:40 (page 99).*

Coffee Break: 10:40 - 11:10

Wednesday, September 11th, 2019, 11:10 - 12:50**Plenary Session Environmental Statistics**

Chair: Roland Fried, Roland Langrock, Room: Audimax

Plenary Session Environmental Statistics: **Peter Filzmoser.** „Correlation analysis for compositional environmental data.“*Wednesday, 11:10-12:00 (page 35).***Heinz-Grohmann Lecture**

Chair: Jürgen Chlumsky, Room: Audimax

Heinz-Grohmann Lecture: **Bernd Fitzenberger.** „Lohnungleichheit in Deutschland: Fakten, Daten, Analysen“.*Wednesday, 12:00-12:50 (page 36).*

Lunch Break: 12:50 - 14:10

Wednesday, September 11th, 2019, 14:10 - 15:50**Housing: Bauen und Wohnen in Deutschland – aktuelle Befunde**

Chair: Ulrike Rockmann, Room: HS7

Göran Kauermann; Michael Windmann; Ralf Münnich. „Neues vom Mietspiegel – Ein Überblick zur aktuellen Diskussion aus Sicht der Statistik“.*Wednesday, 14:10-14:35 (page 100).*

Ingo Wagner. „Ein Bild sagt mehr als 1000 Worte – interaktive Gemeindekarte(n) zur Bautätigkeit in Deutschland“.

Wednesday, 14:35-15:00 (page 100).

Cordula Schöneich; Markus Teske. „Entwicklung von regionalisierten Häuserpreisindizes in der amtlichen Statistik“.

Wednesday, 15:00-15:25 (page 101).

Jens Dechent. „Energieverwendung in Neubau und Bestand – Gibt es Unterschiede bei den eingesetzten Energien in neuen Wohnbauten im Vergleich zum Wohnungsbestand?“

Wednesday, 15:25-15:50 (page 102).

Regional Statistics: Bevölkerung: Gesundheit und Prognose

Chair: Lothar Eichhorn, Room: HS6

Anja Sonnenburg. „Die regionale Pflegebedürftigkeit in Deutschland: Status Quo und künftige Disparitäten“.

Wednesday, 14:10-14:35 (page 103).

Simon Buhl; **Ann-Kathrin Richter.** „Räumliche Verteilung der Arztpraxen in Nordrhein-Westfalen – Analyse auf der Basis des georeferenzierten statistischen Unternehmensregisters“.

Wednesday, 14:35-15:00 (page 103).

Matthias Dörner; Peter Haller. „Not coming in today – Regional firm productivity differentials and the epidemiology of influenza“.

Wednesday, 15:00-15:25 (page 104).

Michael Kalinowski; Gerd Zika; Frederik Parton; Doris Söhnlein. „Die regionalisierte QuBe-Bevölkerungsprojektion“.

Wednesday, 15:25-15:50 (page 105).

National Accounts, Welfare Measurement

Chair: Albert Braakmann, Room: C10

Stefan Hauf; Tanja Mucha. „Ergebnisse der VGR-Revision 2019 in neuem Gewand – das VGR-Dashboard“.

Wednesday, 14:10-14:35 (page 106).

Xaver Dickopf; Christian Janz; Tanja Mucha. „BIP Flash t+10 ? Eine Machbarkeitsstudie zur beschleunigten Schätzung des Bruttoinlandsprodukts“.

Wednesday, 14:35-15:00 (page 106).

Markus Ahlborn; Thorsten Haug. „Globalisierung in den VGR: Erste Ergebnisse von Pilotuntersuchungen auf europäischer Ebene“.

Wednesday, 15:00-15:25 (page 107).

Ayesha Tantriana; Sri Hartini Rachmad. „Understanding Child Deprivation in Indonesia: The Multiple Overlapping Deprivation Analysis (MODA) Approach“.
Wednesday, 15:25-15:50 (page 108).

Computational Statistics and Data Analysis: Interpretable Machine Learning

Chair: Bernd Bischl, Room: HS5

Przemyslaw Biecek. „Explanation, Exploration and Debugging of Predictive Machine Learning Models“.
Wednesday, 14:10-15:00 (page 109).

Jens Mehrhoff. „Rage Against The Machine – The limits of machine learning for automatic product classification“.
Wednesday, 15:00-15:25 (page 109).

Jana Marešková; **Winfried Pohlmeier.** „How well can Noncognitive Skills Predict Unemployment? A Machine Learning Approach“.
Wednesday, 15:25-15:50 (page 110).

Statistical Theory and Methods (Session 1)

Chair: Matei Demetrescu, Room: HS4

Nazarii Salish. „Functional Time Series Analysis“.
Wednesday, 14:10-15:00 (page 111).

Pierre Lafaye De Micheaux; **Pavlo Mozharovskyi;** Myriam Vimond. „Depth for curve data and applications“.
Wednesday, 15:00-15:25 (page 111).

Matei Demetrescu; Nazarii Salish. „Testing the predictive accuracy of functional forecasts“.
Wednesday, 15:25-15:50 (page 112).

FDZ: Neuere methodische Entwicklungen in den FDZ der statistischen Ämter des Bundes und der Länder

Chair: Hans-Ullrich Mühlenfeld, Room: C9

Hans-Ullrich Mühlenfeld; Sven Bocker. „Längsschnittverknüpfung von MZ-Panel-Daten“.
Wednesday, 14:10-14:35 (page 113).

Janina Loske; Tobias Wolfanger. „Entwicklung Synthetischer Datenstrukturfiles in den FDZ“.
Wednesday, 14:35-15:00 (page 113).

Ramona Voshage; Michael Rößner. „Der EU-Unternehmensbegriff – was heißt das für die Firmendaten der amtlichen Statistik in Deutschland (AFiD-Daten) im FDZ“.
Wednesday, 15:00-15:25 (page 114).

Coffee Break: 15:50 - 16:20

Wednesday, September 11th, 2019, 16:20 - 18:00

Housing: Leben am Rande? – Wohnsituation spezifischer Bevölkerungsgruppen

Chair: Sascha Krieger, Room: HS7

Claudia Korn. „Was kann die Statistik der Bundesagentur für Arbeit (BA) zum Thema Wohnen aussagen? Wie wohnen SGB-II-Leistungsberechtigte?“

Wednesday, 16:20-16:45 (page 115).

Christiane Tentscher. „Wie wohnen geflüchtete Menschen in Deutschland, die auf Leistungen der Grundsicherung (SGB II) angewiesen sind? – Eine Analyse struktureller und regionaler Besonderheiten.“

Wednesday, 16:45-17:10 (page 115).

Guido Oemmelen. „Sinus-Geo-Milieus[®] Migranten“.

Wednesday, 17:10-17:35 (page 116).

Labour Market and Social Security

Chair: Holger Meinken, Room: C10

Sandra Dummert; Ute Leber; **Barbara Schwengler.** „Unbesetzte Ausbildungsplätze – Welche regionalen und betrieblichen Gründe spielen eine Rolle?“

Wednesday, 16:20-16:45 (page 118).

Michael Hartmann. „Berufliche Mobilität von Arbeitslosen“.

Wednesday, 16:45-17:10 (page 118).

Thomas Rothe. „Beschäftigungsstabilität nach Langzeitarbeitslosigkeit“.

Wednesday, 17:10-17:35 (page 119).

Bossler Mario; Johann Fuchs; **Alexander Kubis;** Lutz Schneider. „Mögliche Auswirkungen des Brexit auf die künftige Zuwanderung nach Deutschland und die Beschäftigungsentwicklung einheimischer Betriebe“.

Wednesday, 17:35-18:00 (page 120).

Computational Statistics and Data Analysis: Time Series

Chair: Roland Fried, Room: HS5

Jann Goschenhofer; Franz Pfister; Kamer Yuksel; Urban Fietzek; Bernd Bischl. „Wearable-based Parkinson's Disease Severity Monitoring using Deep Learning“.

Wednesday, 16:20-16:45 (page 121).

Dominik Ballreich. „Better time series forecasting using expert knowledge: The Maximum Entropy Approach“.

Wednesday, 16:45-17:10 (page 121).

Alexander Dürre; Roland Fried. „Detecting changes in the dependence structure of a time series“.

Wednesday, 17:10-17:35 (page 122).

Hermann Singer. „Continuous-Discrete Filtering using the Zakai Equation: Smooth Likelihood Surface“.

Wednesday, 17:35-18:00 (page 122).

Statistical Theory and Methods (Session 2)

Chair: Tobias Hartl, Room: HS4

Timo Dimitriadis; Andrew Patton; Patrick Schmidt. „Evaluating the Rationality of Mode Forecasts“.

Wednesday, 16:20-16:45 (page 123).

Simon Wingert; Philipp Sibbertsen. „Local Whittle Wavelet Estimation for Locally Stationary Processes“.

Wednesday, 16:45-17:10 (page 123).

Michelle Voges; Philipp Sibbertsen. „Fractional cointegration in EMU government bond markets“.

Wednesday, 17:10-17:35 (page 124).

Tobias Hartl; Rolf Tschernig; Enzo Weber. „Fractional trends in unobserved components models“.

Wednesday, 17:35-18:00 (page 124).

Methodology of Statistical Surveys: Sampling and Estimation

Chair: Anne Konrad, Room: HS6

Thomas Zimmermann; Ralf Droßard; Simone Scharfe. „Überlegungen zur Stichprobenmethodik der neuen digitalen Verdiensterhebung“.

Wednesday, 16:20-16:45 (page 125).

Stefan Zins; Birgit Horneffer. „Varianzschätzung für den EU-Indikator AROPE“.

Wednesday, 16:45-17:10 (page 125).

Anna-Lena Wölwer; Ralf Münnich. „Unterjährige Schätzung der Arbeitskräfteerhebung im neuen System der Haushaltsstatistiken“.

Wednesday, 17:10-17:35 (page 126).

Anne Konrad; Yves Berger. „Ein multivariater GREG Schätzer zur Schätzung von Veränderungen über die Zeit in Rotationsstichproben“.

Wednesday, 17:35-18:00 (page 126).

VDSSt: Leben und Wohnen in Stadt und Region, Strukturen und Entwicklungschancen

Chair: Ansgar Schmitz-Veltin, Room: C9

Dörte Nitt-Drießelmann. „Pendeln und Wohnkosten sparen: Preisgefälle im Stadt-/Umlankontinuum und deren Bewertung im Wohnatlas 2019“.

Wednesday, 16:20-16:45 (page 128).

Tobias Held; Attina Mäding. „Das Ende der Reurbanisierung? Aktuelle Trends auf dem Wohnungsmarkt der Region Stuttgart“.

Wednesday, 16:45-17:10 (page 128).

Jürgen Gödecke-Stellmann; Theresa Lauerbach. „Entwicklungsdynamiken in Großstadtreionen. Eine Betrachtung ausgewählter Indikatoren auf Basis der Innerstädtischen Raumbewachung und der Laufenden Raumbewachung“.

Wednesday, 17:10-17:35 (page 128).

Jan Goebel. „Soziokulturelle Unterschiede in Großstädten und deren Umland – Ein Entwicklungsvergleich auf der Basis des SOEP“.

Wednesday, 17:35-18:00 (page 129).

Other events on Wednesday

18:00 : General Assembly Meeting DStatG ▷ HS5

19:30 : Get-together at the restaurant „Historischer Keller“

Thursday, September 12th, 2019, 09:00 - 10:40

Business and Market Statistics (Session 1)

Chair: Andreas Kladroba, Room: C10

Simon Rommelspacher; Christian Salwiczek. „Intensive Profiling – Überblick über einen neuen methodischen Ansatz in der amtlichen Unternehmensstatistik“.

Thursday, 9:00-9:25 (page 130).

Maren Koehlmann. „Anwendung einer robusten Regression zur Schließung von Datenlücken am Beispiel der Strukturhebung Energie“.

Thursday, 9:25-9:50 (page 131).

Gloria Gheno. „A new non-monotone link function for marketing analysis“.

Thursday, 9:50-10:15 (page 132).

Malte Schierholz; Tobias Büttner. „Welche Arbeitsagentur ist „auffällig“? Zum Vergleich von Performance-Indikatoren“.

Thursday, 10:15-10:40 (page 132).

Statistical Theory and Methods (Session 3)

Chair: Markus Fritsch, Room: HS5

Kira Alhorn; Holger Dette; Kirsten Schorning. „Model averaging in non-nested models with an application to optimal designs“.

Thursday, 9:00-9:25 (page 134).

Jan Pablo Burgard; **Patricia Dörr.** „Survey-weighting of Linear Mixed Models under Box-Cox and Dual Transformations“.

Thursday, 9:25-9:50 (page 134).

Andrew Adrian Yu Pua; Markus Fritsch; **Joachim Schnurbus.** „Large sample properties of an IV estimator based on the Ahn and Schmidt moment conditions“.

Thursday, 9:50-10:15 (page 135).

Andrew Adrian Yu Pua; **Markus Fritsch;** Joachim Schnurbus. „Practical aspects of using nonlinear moment conditions in linear dynamic panel data models“.

Thursday, 10:15-10:40 (page 135).

Methodology of Statistical Surveys: Machine Learning in Official Statistics

Chair: Thomas Riede, Room: HS6

Joerg Feuerhake. „Maschinelles Lernen in den Handwerksstatistiken“.

Thursday, 9:00-9:25 (page 136).

Florian Dumpert. „Maschinelles Lernen zur Imputation: Eine kritische Betrachtung“.

Thursday, 9:25-9:50 (page 136).

Martin Beck; Paul Mätzig. „Einführung einer automatisierten Plausibilitätskontrolle in der neuen digitalen Verdiensterhebung (NVE)“.

Thursday, 9:50-10:15 (page 137).

Statistical Literacy

Chair: Katharina Schüller, Room: HS7

Tanja Ihden. „„Es liegt nicht immer nur am Empfänger“ – Kommunikation statistischer Ergebnisse durch Statistiker“.

Thursday, 9:00-9:25 (page 138).

Katharina Schüller. „Data Literacy: Schlüsselkompetenz des 21. Jahrhunderts“.

Thursday, 9:25-9:50 (page 138).

Tim Hochgürtel. „Separierung von demografischen und verhaltensbedingten Einflüssen auf die Entwicklung der Lebensformen“.

Thursday, 9:50-10:15 (page 139).

Statistics in Finance: Statistical Learning in Finance

Chair: Yarema Okhrin, Room: HS4

Jozef Barunik; Michael Ellington. „Asset Pricing using Time-Frequency Dependent Network Centrality“.

Thursday, 9:00-9:50 (page 141).

Bastian Gribisch; Tobias Eckernkemper. „Intraday Conditional Value at Risk: A Periodic Mixed-Frequency GAS Approach“.

Thursday, 9:50-10:15 (page 141).

Christian Mücher. „Non-Parametric Prediction of Financial Risk using Artificial Neural Networks. Does High Frequency Information Help?“

Thursday, 10:15-10:40 (page 142).

DGD: Cause of death statistic: Methodological challenges and empirical findings (Session 1)

Chair: Eva Kibele, Room: C9

Olaf Eckert. „Das Kodiersystem Iris/MUSE – Basis der multi- und unikausalen Todesursachenstatistik in Deutschland“.

Thursday, 9:00-9:25 (page 145).

Annelene Wengler; Heike Gruhl; Dietrich Plaß; Janko Leddin; **Elena von der Lippe**. „Ungültige Codes in der Todesursachenstatistik, Definition und regionale Unterschiede“.
Thursday, 9:25-9:50 (page 146).

Coffee Break: 10:40 - 11:10

Thursday, September 12th, 2019, 11:10 - 12:50

Plenary Session Housing

Chair: Ulrike Rockmann, Room: Audimax

Plenary Session Housing: **Marietta Haffner**. „Housing market developments and housing affordability in the European Union.“
Thursday, 11:10-12:00 (page 36).

Gumbel Lecture

Chair: Yarema Okhrin, Room: Audimax

Gumbel Lecture: **Axel Bücher**. „On the block maxima method in extreme value statistics“.
Thursday, 12:00-12:50 (page 37).

Lunch Break: 12:50 - 14:10

Thursday, September 12th, 2019, 14:10 - 15:50

Joint DStatG & SIS – Survey Statistics (Session 1)

Chair: Monica Pratesi, Room: HS6

Alessandra Petrucci; Emilia Rocco. „A proposal to assess the representativeness of non-probability surveys“.
Thursday, 14:10-14:35 (page 143).

Natalie Shlomo; Clare Saunders. „Techniques for randomising protest survey samples to facilitate comparison with nationally representative samples“.
Thursday, 14:35-15:00 (page 143).

Ralf Münnich; Simon Lenau. „Bias correction for non-probability samples“.
Thursday, 15:00-15:25 (page 144).

„General Discussion: Non-Probability Sampling Methods“.
Thursday, 15:25-15:50.

Business and Market Statistics (Session 2)

Chair: Solveigh Jäger, Room: C10

Luis-Federico Flores; Luisa Baumgärtner. „Ergebnisse eines Eurostat-Projektes zur Modernisierung der europäischen Produktionsstatistik“.

Thursday, 14:10-14:35 (page 147).

Angelika Becker. „Zirkuläre Wirtschaft – Herausforderungen für Industrie und Statistik“.

Thursday, 14:35-15:00 (page 148).

Muhamed Kudic; **Katharina Friz**; Andreas Kladroba; Tobias Buchmann; **Patrick Wolf.** „Netzwerk- und Outputmessung – Indikatorik für transformative Technologiefelder“.

Thursday, 15:00-15:25 (page 148).

Andreas Kladroba. „Wie geht man mit Lücken in Eurostat-Daten um? Vorschläge am Beispiel der FuE-Erhebung.“

Thursday, 15:25-15:50 (page 149).

Statistics in Finance (Session 2)

Chair: Yarema Okhrin, Room: HS4

Janis Becker; Christian Leschinski. „The Bias of Realized Volatility“.

Thursday, 14:10-14:35 (page 150).

Holger Dette; Vasyl Golosnoy; **Janosch Kellermann.** „How to correct the intraday periodicity bias in realized volatility measures“.

Thursday, 14:35-15:00 (page 150).

Eugen Ivanov; **Yarema Okhrin.** „Matrixvariate Factor Model for Realized Covariances“.

Thursday, 15:00-15:25 (page 151).

Andreas Masuhr. „Big in Japan: Global Volatility Transmission between Assets and Trading Places“.

Thursday, 15:25-15:50 (page 151).

DGD: Cause of death statistic: Methodological challenges and empirical findings (Session 2)

Chair: Elena von der Lippe; Enno Nowossadeck, Room: C9

Susanne Stolpe; Andreas Stang. „Regionale Unterschiede in der Verwendung von nicht-informativen ICD-Kodierungen bei kardiovaskulären Todesfällen und deren Auswirkung auf die Vergleichbarkeit von Mortalitätsraten für kardiovaskuläre Erkrankungen“.

Thursday, 14:10-14:35 (page 152).

Sebastian Franke. „Ein Vergleich der Ungleichheit anhand verlorener Lebensjahre zwischen Ost und West Deutschland zwischen 2003 und 2015“.

Thursday, 14:35-15:00 (page 153).

Holger Leerhoff. „Das Suizidgeschehen in Deutschland“.

Thursday, 15:00-15:25 (page 154).

Regional Statistics: Sinus-Milieus, Ausländer, Diverses

Chair: Lothar Eichhorn, Room: HS7

Britta Stöver. „Die wirtschaftliche Bedeutung von Hochschulen für ihre Region“.

Thursday, 14:10-14:35 (page 155).

Guido Oemmelen. „Daten zur Identifikation ethnisch diversifizierter Sozialräume“.

Thursday, 14:35-15:00 (page 155).

Madeline Kaupert. „Daseinsvorsorge und interkommunale Kooperation in kleineren Städten und Gemeinden“.

Thursday, 15:00-15:25 (page 156).

Statistics in Natural Science and Technology (Session 1)

Chair: Ansgar Steland, Room: HS5

Rainer von Sachs; Marion Rebecca; Govaerts Bernadette; Lederer Johannes. „Selecting variables via clustering within regularized prediction in high dimensions“.

Thursday, 14:10-15:00 (page 157).

Monika Bours; Ansgar Steland. „Asymptotics for High-Dimensional Covariance Matrices of Factor Models“.

Thursday, 15:00-15:25 (page 157).

Nils Mause; Ansgar Steland. „Detecting Changes in the Second Moment Structure of High-Dimensional Sensor-Type Data in a K-Sample Setting“.

Thursday, 15:25-15:50 (page 158).

Coffee Break: 15:50 - 16:20

Thursday, September 12th, 2019, 16:00 - 16:20

Springer Best-Poster-AwardSpringer booth

Thursday, September 12th, 2019, 16:20 - 18:00

Data Journalism (Session 1)

Chair: Jörg Rahnenführer, Room: HS7

Thomas Weißbrodt; Stefan Werth. „BA-Statistik für Datenjournalismus nutzen“.*Thursday, 16:20-16:45 (page 160).***Lars Koppers; Sarah Kohler; Meik Bittkowski; Nikolai Promies; Volker Stollorz; Markus Lehmkuhl.** „Entwicklung von Methoden und Tools für eine datengestützte Wissenschaftskommunikation“.*Thursday, 16:45-17:10 (page 160).***Jonas Rieger; Carsten Jentsch; Jörg Rahnenführer.** „Quantifizierung der Stabilität der Latent Dirichlet Allocation mithilfe von Clustering auf wiederholten Durchläufen“.*Thursday, 17:10-17:35 (page 161).*

Bayesian Analysis in Natural and Social Science (Session 1)

Chair: Olha Bodnar, Room: C10

Richard Nickl; Jakob Söhl. „Bernstein-von Mises Theorems for Compound Poisson Processes“.*Thursday, 16:20-17:10 (page 162).***Nestor Parolya; Taras Bodnar; Wolfgang Schmid; David Bauder.** „Bayesian Inference of the Multi-Period Optimal Portfolio for an Exponential Utility“.*Thursday, 17:10-17:35 (page 162).***Taras Bodnar; David Bauder; Nestor Parolya; Wolfgang Schmid.** „Bayesian mean-variance analysis: Optimal portfolio selection under parameter uncertainty“.*Thursday, 17:35-18:00 (page 163).*

Joint DStatG & SIS – Survey Statistics (Session 2)

Chair: Ralf Münnich, Room: HS6

Monica Pratesi; Luciana Quattrociocchi; Gaia Bertarelli; Alessandro Gemignani; Caterina Giusti. „Spatial distribution of multidimensional educational poverty in Italy using small area estimation“.*Thursday, 16:20-17:10 (page 164).*

Jan Pablo Burgard. „Measurement Errors in Small Area Estimation“.
Thursday, 17:10-18:00 (page 164).

Statistics in Finance (Session 3)

Chair: Theo Berger, Room: HS4

Theo Berger. „Volatility spillover along the supply chains: A network analysis on economic links“.
Thursday, 16:20-16:45 (page 165).

Markus Haas. „Conditional skewness and kurtosis of aggregated normal mixture GARCH returns with application to Value-at-Risk and Expected Shortfall forecasting“.
Thursday, 16:45-17:10 (page 165).

Oliver Old. „Asymptotic Properties of GARCH Models in the Presence of Time Varying Unconditional Variance. A Simulation Study“.
Thursday, 17:10-17:35 (page 166).

Taras Bodnar; **Solomiia Dmytriv**; Yarema Okhrin; Nestor Parolya; Wolfgang Schmid. „New Approaches in Testing High-Dimensional EU Portfolios“.
Thursday, 17:35-18:00 (page 166).

Statistics in Natural Science and Technology (Session 2)

Chair: Rainer von Sachs, Room: HS5

Matthias Gärtner; Sevil Duvarci; Jochen Roeper; **Schneider, Gaby.** „Statistical analysis of joint pausing in parallel spike trains“.
Thursday, 16:20-16:45 (page 167).

Sven Knoth. „Improving the ARL profile and the accuracy of its calculation for Poisson EWMA charts“.
Thursday, 16:45-17:10 (page 167).

Markus Fritsch; Harry Haupt; Joachim Schnurbus; Philipp Sibbertsen; **Kai Rouven Wenger.** „Time series diagnostics and forecasting of rainfall data“.
Thursday, 17:10-17:35 (page 168).

Ivan Semeniuk; Wolfgang Schmid; Yarema Okhrin. „Monitoring image processes assuming spatially correlated pixels“.
Thursday, 17:35-18:00 (page 168).

Other events on Thursday

19:30 : Conference Dinner

Friday, September 13th, 2019, 09:00 - 10:40

MAKSWELL EU: Use of big data for measuring poverty and well-being

Chair: Fabio Bacchini, Room: HS5

Monica Pratesi; Gaia Bertarelli; Luigi Biggeri; Caterina Giusti; Stefano Marchetti; Francesco Schirripa-Spagnolo. „Small area poverty indicators adjusted using local price indexes“.
Friday, 9:00-10:25 (page 170).

Paul Smith; Nikos Tzavidis; Angela Luna; Jessica Steele; Kristine Nilsen. „Alternative spatial data sources for small area estimation in developing countries“.
Friday, 9:25-9:50 (page 170).

Ralf Münnich. „Measuring well-being at local level using big data“.
Friday, 9:50-10:15 (page 171).

Natalie Shlomo; Martin Karlberg. „Discussion“.
Friday, 10:15-10:40.

Data Journalism (Session 2)

Chair: Lars Koppers, Room: HS7

Susanne Kirschstein-Barczewski. „Welche Chancen und Hürden bietet die Digitalisierung im schulischen Bildungskontext? – Ein Datenjournalismusprojekt“.
Friday, 9:00-9:25 (page 172).

Paul Klammer; Holger Wormer. „Daten-Kompetenz für Journalisten – das Dortmunder Modell“.
Friday, 9:25-9:50 (page 173).

Hans-Joachim Mittag; Katharina Schüller. „Wozu brauchen Datenjournalisten „Statistical Literacy“ und welchen Mehrwert können interaktive Visualisierungen für Online-Publikationen haben?“
Friday, 9:50-10:15 (page 173).

Bayesian Analysis in Natural and Social Science (Session 2)

Chair: Nestor Parolya, Room: C10

Göran Kauermann; **Benjamin Sischka.** „Bayesian and Spline based Approaches for (EM based) Graphon Estimation“.
Friday, 9:00-9:25 (page 175).

Olha Bodnar. „Non-Informative Bayesian Inference for Heterogeneity in a Generalized Marginal Random Effects Meta-Analysis“.
Friday, 9:25-10:15 (page 175).

Joseph Sakshaug; Arkadiusz Wisniowski; Diego Perez-Ruiz; Annelies Blom. „Bayesian Integration of Probability and Non-Probability Samples: Implications for Survey Estimation and Survey Costs“.
Friday, 10:15-10:40 (page 176).

Statistics in Finance (Session 4)

Chair: Markus Haas, Room: HS4

Fabian Mies. „Rate-efficient estimation of the BG index of a Lévy process“.
Friday, 9:00-9:25 (page 177).

Timo Dimitriadis; Julie Schnaitmann. „Encompassing Tests for Higher-Order Elicitable Functionals“.
Friday, 9:25-9:50 (page 177).

Manuel Schmid; Ostap Okhrin; Michael Rockinger. „Stochastic Volatility Jump Diffusion Models: Moments, Estimation and Simulation“.
Friday, 9:50-10:15 (page 178).

Taras Bodnar; **Dmytro Ivasiuk;** Nestor Parolya; Wolfgang Schmid. „An approximate solution for the power utility optimization under predictable returns“.
Friday, 10:15-10:40 (page 178).

Methodology of Statistical Surveys: Miscellaneous

Chair: Thomas Zimmermann, Room: HS6

Sabine Zinn; **Timo Gnambs.** „Using Bayesian Additive Regression Trees for Modelling Attrition and Unit Non-Response in Longitudinal Surveys“.
Friday, 9:00-9:25 (page 179).

Hannah Kiiver; Frank Espelage. „Longitudinal weights and attrition in the EU-Labour Force Survey“.
Friday, 9:25-9:50 (page 179).

Christian Joebges. „From Field Survey to Land Use Statistics – Supporting the LUCAS project“.
Friday, 9:50-10:15 (page 180).

Sigbert Klinke; Jasmin Schlax. „How to Find Him or Her Again?“
Friday, 10:15-10:40 (page 181).

Coffee Break: 10:40 - 11:10

Friday, September 13th, 2019, 11:10 - 12:50

MAKSWELL EU: MAKING Sustainable development and WELL-being frameworks work for policy

Chair: Monica Pratesi, Room: HS5

Fabio Bacchini; Maria Grazia Calza. „Introduction to the project MAKSWELL“.

Friday, 11:10-11:35 (page 182).

Giorgio Alleva; Martin Karlberg; Natalie Shlomo; Gaby Umbach. „Round Table Discussion“.

Friday, 10:35-12:50.

Friday, September 13th, 2019, 11:10 - 12:30

Plenary Session Data Journalism

Chair: Jörg Rahnenführer, Room: Audimax

Plenary Session Data Journalism: **Marcel Pauly.** „What data journalism is and how cooperation with scientists and statisticians can work“.

Friday, 11:10-12:00 (page 37).

Panel Discussion: Data Journalism

Chair: Jörg Rahnenführer, Room: Audimax

Panel Discussion Data Journalism.

Friday, 12:00-12:30 (page 38).

Other events on Friday

9:00 : R-Tutorial: Small Area Estimation with R ▷ C360

1 Prospective microsimulation of pensions in European Member States

Gijs Dekkers; Van Den Bosch Karel

Federal Planning Bureau, Belgium
gd@plan.be

SECTION: Microsimulations – Applications and Methods

Many public or semi-public research agencies and ministries in EU member states use microsimulation models (MSMs) to make projections of future pensions. As will become clear below, for good reasons, most of these models employ dynamic ageing techniques, though some use a static ageing method. The aim of this chapter is to discuss these models, many of which were not mentioned in the overview by Li and O'Donoghue (2013). Though different in many ways, they share common characteristics, of which their use of administrative data as the primary database is the most remarkable one. Survey data play a supplementary role, to impute missing data, or to estimate behavioural equations. Furthermore there are extensive ad-hoc collaborative ties between various teams, partly driven by various EU-funded research projects. As a result, even though the ambition of a network for dynamic microsimulation has not been realized yet (Dekkers and Zaidi, 2011), collaboration between various teams is now more intense than ever.

2 Correlation analysis for compositional environmental data

Peter Filzmoser

Vienna University of Technology, Austria
P.Filzmoser@tuwien.ac.at

SECTION: Environmental Statistics

Many environmental data sets consist of concentration values of different chemical elements or compounds, and already their unit (mg/kg, ppm, etc.) reflects that the information refers to a certain total. Analyzing „relative“ information is the purpose of compositional data analysis, and the most prominent methods in this area are based on the so-called log-ratio methodology. Particularly when analyzing the correlation between pairs of variables it is known that the traditional non-compositional methods can result in spurious correlations. This is caused by a tendency towards negative covariances, and thus the correlations lose their predicative value.

There are different possibilities to avoid this problem. We will consider so-called symmetric pivot coordinates. Their purpose is to build an orthonormal coordinate system, expressing all relative information in terms of aggregated log-ratios, where the first two coordinates refer to all relative information of the two variables of interest. Since log-ratios could become unstable in case of data quality problems, we propose an extension to weighted symmetric pivot coordinates, where appropriate weights can be assigned to the variables. The usefulness of these methods will be demonstrated with simulated data and data sets from geochemistry.

3 Lohnungleichheit in Deutschland: Fakten, Daten, Analysen

Bernd Fitzenberger

HU Berlin, Deutschland
fitzenbb@hu-berlin.de

SECTION: Heinz-Grohmann Lecture

In Deutschland ist in den letzten Jahrzehnten ein deutlicher Anstieg der Lohnungleichheit zu beobachten, der sich bis zum Beginn der 1990er Jahre auf den oberen Bereich der Lohnverteilung beschränkte und der sich bis 2010 kontinuierlich sowohl im oberen als auch im unteren Bereich der Lohnverteilung fortsetzte (Dustmann et al. 2009, Antonczyk et al. 2010a, 2010b, Card et al. 2013). Nach 2011 ist eine „Trendwende“ zu konstatieren, da der Anstieg der Lohnungleichheit zum Stillstand kam (Möller 2016, Fitzenberger und Seidlitz 2019). Der Anstieg der Lohnungleichheit ging mit dem Anwachsen des Niedriglohnsektors sowie einer Stagnation oder gar einem Rückgang der Reallöhne am unteren Ende der Lohnverteilung einher. Im Rahmen der Grohmann-Vorlesung soll zunächst eine zusammenfassende systematische Darstellung und Diskussion der Entwicklung der Lohnungleichheit in Deutschland auf Basis der Ergebnisse von einschlägigen wissenschaftlichen Studien und unter Rückgriff auf aktuelle eigene Arbeiten erfolgen. Ein Schwerpunkt liegt auf dem Vergleich verschiedener Datensätze, der Diskussion von möglichen Messfehlern und der Darstellung und Anwendung methodischer Verfahren zur Analyse der Lohnungleichheit.

4 Housing market developments and housing affordability in the European Union

Marietta Haffner

Delft University of Technology, Niederlande
m.e.a.haffner@tudelft.nl

SECTION: Housing

This presentation provides a brief overview of the housing situation in the European Union based on the Eurostat data from the European Union Statistics on Income and Living Conditions (EU-SILC). A history in a nutshell shows that (national) housing policies have shifted towards the market in different ways since the 1980s. What this implies for the present state of the art in the affordability of housing is the main focus of the presentation. Housing affordability is defined as outlays for housing that consumers of housing services (rent, mortgage costs, energy costs, etc.) are making. The housing affordability outcomes for different definitions of housing affordability, such as consumer perceptions and the housing cost overburden ratio, are discussed and compared across different groups of consumers. Furthermore, these are linked to some developments in housing policy and contexts affecting housing markets. The outcomes illustrate that the population „at risk of poverty“ and „tenants paying market rent“ are most at risk of encountering housing affordability problems. More conceptually, making explicit the justifications for applying certain approaches to housing affordability must be considered a must in the design of housing policies.

5 On the block maxima method in extreme value statistics

Axel Bücher

Heinrich-Heine-Universität Düsseldorf, Deutschland
axel.buecher@hhu.de

SECTION: Gumbel-Vorlesung

For the analysis of extreme values, two fundamental approaches can be distinguished. First, the peaks-over-threshold method consists of extracting those values from the observation period which exceed a high threshold. To model such threshold excesses, asymptotic theory suggests the use of the Generalized Pareto distribution. Second, the block maxima method, popularized by Emil Gumbel in 1958, consists of dividing the observation period into a sequence of non-overlapping intervals and restricting attention to the largest observation in each time interval. Thanks to the extremal types theorem, the probability distribution of such block maxima is approximately Generalized Extreme-Value (GEV), which may then be fitted to the sample of block maxima. In this talk, we review some recent theoretical results on the block maxima method, and compare them to existing results for the peaks-over-threshold approach. We further propose some methodological improvements on the block maxima method.

6 What data journalism is and how cooperation with scientists and statisticians can work

Marcel Pauly

Spiegel Online, Deutschland
marcel.pauly@spiegel.de

SECTION: Data Journalism

Data journalists use datasets and statistical methods to find and tell stories. Marcel Pauly, head of Der Spiegel's data unit, gives an introduction to the topic: How do data journalists work? Why are their skills important for media organizations? How do data teams organize themselves in different news outlets?

After this general introduction the talk will focus on similarities and differences between data journalism and science as well as statistics. Data gathered and collected by scientists can be the basis for journalistic investigations on topics like climate and weather phenomena, road fatalities, political populism or court decisions. A better understanding of each other's needs and working conditions will help to improve cooperation.

7 Panel Discussion Data Journalism

Jörg Rahnenführer

Technische Universität Dortmund, Deutschland
rahnenuhruhr@statistik.tu-dortmund.de

SECTION: Data Journalism

Im Datenjournalismus verbindet sich klassischer Journalismus mit Daten, Statistik und Informatik. Die Open Data Bewegung ermöglicht es zunehmend, dass Unmengen von Datensätzen zur Analyse für die Allgemeinheit zur Verfügung stehen. Datenjournalisten arbeiten mit statistischen Verfahren und Analyse-tools, um interessante Geschichten auf Grundlage u.a. solcher Daten zu erzählen, oft auch in interaktiver Form. Anknüpfend an den Plenarvortrag werden in der Abschlussdiskussion Anknüpfungspunkte zwischen Datenjournalismus und Statistik im Spannungsfeld von aktueller Berichterstattung, Verständlichkeit in der Darstellung, und wissenschaftlicher Genauigkeit besprochen. Dazu werden TeilnehmerInnen der Konferenz aus dem Datenjournalismus und aus der Statistik kurz Stellung beziehen, bevor die Diskussion für Fragen und Bemerkungen aus dem Publikum geöffnet wird.

8 The International Journal of Microsimulation. Eine bibliometrische Netzwerkanalyse über Akteure und Anwendungsfelder

Marc Hannappel; Ardian Canolli

Universität Koblenz-Landau, Germany
MarcHannappel@uni-koblenz.de

SECTION: Microsimulations – Applications and Methods

Es ist nun mehr als sechzig Jahre her, dass Guy Orcutt seine Vision einer damals neuen Simulationsmethode veröffentlichte. Es dauerte aber genau fünfzig weitere Jahre, bis seine Vision in Form eines eigens für Mikrosimulationen vorgesehenen Publikationsorgans institutionalisiert wurden. Die Gründung des IJM ist nun gut elf Jahre her. Das Jubiläumsjahr wurde auch bereits genutzt, um die Entwicklungsgeschichte der Zeitschrift reflexiv zu bilanzieren. Die Ergebnisse dieser Bilanzen fallen dabei durchaus gemischt aus.

Wir möchten diese Betrachtungen mit einem Blick von außen ergänzen. Mit Hilfe einer bibliometrischen Netzwerkanalyse aller Artikel und Beiträge des IJM möchten wir uns ein Bild über den aktuellen Stand der Mikrosimulation sowie über dessen Genese innerhalb der Zeitschrift machen, wobei wir die Entwicklung innerhalb der Zeitschrift durchaus als einen Indikator für den Stand der Mikrosimulation innerhalb der Sozialwissenschaft insgesamt betrachten.

Wir gehen dabei folgenden Fragen nach:

- (1) Wer sind die entscheidenden Akteure innerhalb des IJM?
- (2) Lassen sich Personencluster mit besonderer Publikationsstärke und Vernetzungsgrad identifizieren?
- (3) Welche thematischen Schwerpunkte lassen sich identifizieren und wie sind diese Themen miteinander verbunden?

9 Mikrosimulation und die Hochrechnung von Mikrodaten – Erfahrungen und Anwendungen eines informationstheoretischen Ansatzes und dem Programmpaket ADJUST

Joachim Merz

Leuphana Universität Lüneburg, Germany
merz@uni.leuphana.de

SECTION: Microsimulations – Applications and Methods

Mikrosimulationsmodelle (MSM) haben sich erfolgreich bewährt, die Wirkungen alternativer Politiken zu quantifizieren. Im wirtschafts- und sozialpolitischen Bereich auf internationaler und nationaler Ebene – und natürlich wenn reale Experimente nicht möglich sind – hilft die MSM-Analyse alternativer Szenarien Effekte aufzudecken, die alleine bei der Analyse typischer Fälle nicht sichtbar werden. Dazu ist es notwendig auf der internationalen, nationalen, regionalen Ebene, gesamtgesellschaftlich wie auch gruppenbezogen (z.B. zunehmend auf der Unternehmerebene), repräsentative Daten und Ergebnisse zu erhalten.

Die Hochrechnung, die Kalibrierung von Mikrodaten ist dafür ein effizientes Mittel, durch (Um-)Gewichtung der Mikroeinheiten (Familien, Haushalte, Betriebe, Organisationen) Stichproben, die nicht im strengen Sinn zufällig sind, für repräsentative Aussagen und Folgerungen zu gewinnen. Somit hat die Hochrechnung generell von Mikrodaten eine über den Einsatz in Mikrosimulationsmodellen hinausgehende weitreichende Bedeutung.

In diesem Vortrag wird – im Rahmen statischer (static aging), dynamischer Mikrosimulationsmodelle als auch für artificial agency-Modelle – ein Hochrechnungsverfahren auf informationstheoretischer Basis vorgestellt. Anhand von internationalen und nationalen Erfahrungen und Anwendungen wird sein erfolgreicher Einsatz mit dem Programmpaket ADJUST nicht nur für die Kalibrierung sondern auch für Sensitivitätsanalysen und weitergehende MSM-Fragestellungen an Beispielen zur Transfer- und Steuerpolitik, zu Analysen zum Armuts- und Reichtumsbericht der Bundesregierung, zur Rentenreform, zum Erfolg von Unternehmen, zur Patientenanalyse, zum Konsumverhalten etc. demonstriert. Eine „realtime“ Hochrechnung von Mikrodaten zeigt abschliessend die effiziente und einfache stand-alone Handhabung von ADJUST im Rahmen unseres neuen MICSIM-4J, einem Mikrosimulationsmodell für die Forschung und Lehre.

10 Mikroanalyse und Georeferenzierung in der amtlichen Statistik

Hanna Brenzel; Markus Zwick

Federal Statistical Office (Destatis), Germany
Hanna.Brenzel@destatis.de

SECTION: Microsimulations – Applications and Methods

Mikroanalyse und -simulation sind für eine zielgenaue Wirtschafts- und Sozialpolitik von fundamentaler Bedeutung, da sie eine Untersuchung der Auswirkungen politischer Entscheidungen für direkt Betroffene ermöglichen. Um das Potential von Mikrosimulationen und ihre diversen Anwendungsfelder in der amtlichen Statistik zu verdeutlichen, werden u.a. die Historie des Themas und gegenwärtige Projekte vorgestellt.

In der amtlichen Statistik ist Mikrosimulation kein neues Thema. Die Grundlagen für die mikrodatenbasierte Analyse schuf der durch die Deutsche Forschungsgemeinschaft (DFG) finanzierte Sonderforschungsbereich 3 „Mikroanalytische Grundlagen der Gesellschaftspolitik“ der Universitäten Frankfurt und Mannheim. Zentrale amtliche Daten waren hier vor allem die Einkommens- und Verbraucherstichprobe sowie der Mikrozensus.

Mit der Novellierung des § 3 (1) Nr. 6 Bundesstatistikgesetz (BStatG) im Sommer 2016 ist es nun u.a. Aufgabe des Statistischen Bundesamts „Entwicklung und Anwendung von Mikrosimulationsmodellen sowie mikroökonomische Analysen durchzuführen“. Zur Operationalisierung des neuen gesetzlichen Auftrages befasst sich das Statistische Bundesamt derzeit mit dem Aufbau eines Methodenzentrums zur Mikrosimulation und -analyse.

Mit der Anpassung des § 10 BStatG im August 2013 wurde der Grundstein für die Georeferenzierung im Statistischen Verbund gelegt: Gemäß der Gesetzesänderung sind neben der Nutzung der sogenannten Blockseite nun auch geografische Gitterzellen als feinste regionale Zuordnung zulässig. Die Erreichung des Ziels der Amtlichen Statistik, Informationen bereitzustellen, die in demokratischen Gesellschaften zur Willensbildung genutzt werden, und die Grundlage für Entscheidungsprozesse darstellen, wird durch die Georeferenzierung statistischer Informationen demnach um eine wesentliche Facette erweitert. Die Geo-

referenzierung schafft es somit den steigenden Informationsbedarf von Politik, Wissenschaft, Wirtschaft und Gesellschaft zu decken und bildet eine neue, erweiterte Grundlage für evidenzbasierte Entscheidungen – nicht zuletzt durch die Anwendung von Mikroanalysen und -simulationen auf kleinräumiger Ebene. Konkret zeigt sich dies in der Beteiligung des DFG-geförderten Projekts „Sektorenübergreifendes kleinräumiges Mikrosimulationsmodell (MikroSim)“ der Universitäten Trier und Duisburg, indem das Statistische Bundesamt Projektpartner ist.

11 Die Anwendung von Mikrosimulationen zur Bevölkerungsfortschreibung

Julian Reinhold; Philip Höcker; Sven Brocker; Rainer Schnell

Universität Duisburg-Essen, Germany
mikrosimulation@uni-due.de

SECTION: Microsimulations – Applications and Methods

Einwohnerzahlen sind für allgemeine Planungs- und Verwaltungsaufgaben von besonderer Bedeutung. Um hierfür auf verlässliche Schätzungen zurückgreifen zu können, kommt der regelmäßig durchgeführten Fortschreibung des Bevölkerungsstandes eine wichtige Rolle zu. Für eine Betrachtung der derzeit verwendeten Methoden zur Bevölkerungsfortschreibung wurden die statistischen Ämter der Europäischen Union und zusätzlich der Schweiz kontaktiert und die eingesetzten Projektionsmethoden erfasst. In den überwiegenden Fällen wird die Kohorten-Komponenten-Methode verwendet. Diese Methode projiziert die Bevölkerungszahl auf der Basis vorliegende Makrodaten über Geburts-, Migrations- und Sterbefälle. Eine alternative Projektionsmethode stellt die Anwendung von Mikrosimulationen dar, bei denen individuelle Personen mit ihren Merkmalen simuliert werden. Dieser Beitrag stellt die Vor- und Nachteile beider Methoden einander vergleichend gegenüber. Dabei werden die Potentiale der Mikrosimulation in der Bevölkerungsvorausberechnung dargestellt. Demnach eignen sich Mikrosimulationsmethoden vor allem zur Einbeziehung weiterer Merkmale, wohingegen Makromethoden eher bei der Prädiktion von weniger differenzierten Aggregatwerten eingesetzt werden können. Insgesamt bieten Mikrosimulationen bisher wenig genutzte Möglichkeiten für die demographische Forschung und die amtliche Statistik.

12 A Structural Model for the Coevolution of Networks and Behavior

Michael Konig

VU Amsterdam, Niederlande
m.d.konig@vu.nl

SECTION: Empirical Economics

This paper introduces a structural model for the coevolution of networks and behavior. The micro-foundation of our model is a network game where agents adjust actions and network links in a stochastic best-response dynamics with a utility function allowing for both network formation externalities and unobserved heterogeneity. We show the network game admits a potential function and the coevolution process converges to a unique stationary distribution characterized by a Gibbs measure. To bypass the evaluation of the intractable normalizing constant in the Gibbs measure, we adopt the Double Metropolis-Hastings algorithm to sample from the posterior distribution of the structural parameters. To illustrate the empirical relevance of our structural model, we apply it to study R&D investment and collaboration decisions in the chemicals and pharmaceutical industry and find a positive knowledge spillover effect. Finally, our structural model provides a tractable framework for a long-run „key player“ analysis. We provide a key player ranking of the firms in our sample according to their impacts on welfare upon exit. The ranking shows that the key players are not necessarily the firms with the highest market shares.

13 Naïve Learning in Social Networks with Random Communication

Ines Lindner; Huang Jia-Ping; Bernd Heidergott

Vrije Universiteit Amsterdam, Netherlands, The
i.d.lindner@vu.nl

SECTION: Empirical Economics

We study social learning in a social network setting where agents receive independent noisy signals about the truth. Agents naïvely update beliefs by repeatedly taking weighted averages of neighbors' opinions. The weights are fixed in the sense of representing average frequency and intensity of social interaction. However, the way people communicate is random such that agents do not update their belief in exactly the same way at every point in time. We show that even if the social network does not privilege any agent in terms of influence, a large society almost always fails to converge to the truth. We conclude that wisdom of crowds is an illusive concept and bares the danger of mistaking consensus for truth.

14 Dynamic interbank network analysis using latent space models

Fernando Linardi (1); Cees Diks (2,3); **Marco van der Leij (2,3,4)**; Iuri Lazier (1)

1: Central Bank of Brazil, Brazil; 2: University of Amsterdam, The Netherlands; 3: Tinbergen Institute, The Netherlands;
4: De Nederlandsche Bank, The Netherlands
m.j.vanderleij@uva.nl

SECTION: Empirical Economics

Longitudinal network data are increasingly available, allowing researchers to model how networks evolve over time and to make inference on their dependence structure. In this paper, a dynamic latent space approach is used to model directed networks of monthly interbank exposures. In this model, each node has an unobserved temporal trajectory in a low-dimensional Euclidean space. Model parameters and latent banks' positions are estimated within a Bayesian framework.

We apply this methodology to analyze two different datasets: the unsecured and the secured (repo) interbank lending networks. We show that the model that incorporates a latent space performs much better than the model in which the probability of a tie depends only on observed characteristics; the latent space model is able to capture some features of the dyadic data such as transitivity that the model without a latent space is not able to.

15 Regularisierte Small Area Modelle zur robusten Schätzung in der Gegenwart unbekannter Messfehler in den Kovariaten

Joscha Krause; Jan Pablo Burgard; Dennis Kreber

Universität Trier, Deutschland
krause@uni-trier.de

SECTION: Methodology of Statistical Surveys

Small Area Estimation umfasst eine Klasse quantitativer Methoden zur stabilen Schätzung regionaler Statistiken auf Basis kleiner Stichproben. Small Area Schätzer erlauben in einer solchen Situation effizientere Schätzungen als direkte Schätzer, welche bei geringer Fallzahl inakzeptabel hohe Stichprobenvarianzen aufweisen. Dabei wird der Effizienzgewinn durch die funktionale Verbindung von Hilfsinformationen verschiedener Regionen in geeigneten Regressionsmodellen (Small Area Modellen) ermöglicht. Je höher der Erklärungsgehalt des Modells ist, desto größer ist auch der Effizienzgewinn des Small Area Schätzers relativ zu direkten Schätzern.

Die funktionale Verbindung von Kovariaten kann jedoch auch suboptimale Ergebnisse liefern. Dies ist insbesondere dann der Fall, wenn die Hilfsinformationen von Messfehlern betroffen sind. In einer solchen Situation muss der angewendete Small Area Schätzer robustifiziert werden, um valide Ergebnisse zu liefern. In der Literatur werden hierfür häufig Verteilungsannahmen hinsichtlich der Messfehler eingeführt, um unter dieser Prämisse robuste Ergebnisse abzuleiten. Derartige Annahmen sind jedoch oftmals unrealistisch, da in der Regel keine Verteilungsinformationen über die Messfehler vorliegen. Folglich müssen alternative Wege zur Robustifizierung von Small Area Schätzern gefunden werden, welche keine Kenntnis der Messfehlerverteilung voraussetzen.

Wir schlagen regularisierte Modellparameterschätzung in Small Area Modellen vor, um robuste Schätzwerte bei Messfehlern in den Kovariaten zu erhalten. Hierfür wird analytisch gezeigt, dass die regularisierte Schätzung der Regressionskoeffizienten äquivalent ist zur robusten Minimierung der Loglikelihood-Funktion in der Gegenwart unbekannter Messfehler. Diese Äquivalenz wird anschließend verwendet, um ein robustes Fay-Herriot Modell herzuleiten. Zusätzlich wird ein modifizierter Jackknife-Ansatz zur Schätzung des Mean Squared Error entwickelt. Die Effektivität der Methodik wird anhand einer Simulationsstudie verdeutlicht. Zusätzlich erfolgt eine empirische Anwendung im Bereich der Armutsmessung.

16 „City Data from LFS and Big Data“ – Schätzung von Indikatoren der Arbeitskräfteerhebung für funktionale städtische Gebiete mit Mobilfunkdaten

Sandra Hadam (1); Nora Würz (2); Ann-Kristin Kreuzmann (2); Timo Schmid (2)

1: Federal Statistical Office of Germany; 2: Freie Universität Berlin
sandra.hadam@destatis.de

SECTION: Methodology of Statistical Surveys

Zuverlässige Kenntnisse über die Indikatoren für die Erwerbsbevölkerung eines Landes sind für eine solide evidenzbasierte Politikgestaltung unerlässlich. So wird beispielsweise die geografische Verteilung der Beschäftigung oder der Erwerbslosenquote herangezogen, um Entscheidungen über die Verteilung von Ressourcen zu treffen. Die Arbeitskräfteerhebung (AKE) ist eine ESS Gemeinschaftsstatistik und in Deutschland in den Mikrozensus integriert und ist im Allgemeinen darauf ausgerichtet, zuverlässige Schätzungen für größere Bereiche wie die nationale oder regionale Ebene zu liefern. Um jedoch politische Empfehlungen in städtischen Gebieten machen zu können, müssen Informationen für diese Gebiete bereitgestellt werden. Eine Möglichkeit, Schätzungen auf räumlich disaggregierten Ebenen abzuleiten, ist die Verwendung von Small Area Methoden und basiert auf der Verfügbarkeit von prädiktiven Hilfsgrößen. Daher werden neben der Verwendung von Informationen der AKE auch alternative Quellen aus passiv erfassten Mobilfunkgeräten verwendet. Die Grundidee für diese Anwendung ist es, anonymisierte und aggregierte Mobilfunkdaten der Deutschen Telekom als Hilfsvariablen zur Schätzung der Erwerbslosenquote für Funktionale Städtische Gebiete zu verwenden. Die Methodik hängt vom Ansatz und der Vorgehensweise von Schmid et al. (2017) ab, welche soziodemographische Indikatoren im Senegal durch die Kombination von Mobilfunkdaten und Survey Daten geschätzt haben. Die Motivation für die Nutzung von Mobilfunkdaten besteht darin, dass sie ohne Unterbrechung erhoben werden und wertvolle Informationen über Zeitpunkt, Ort und Intensität aggregierter mobiler Aktivitäten enthalten. Da wir aggregierte Mobilfunkdaten für Deutschland nutzten, können wir die Erwerbslosenquote auf räumlich kleinräumiger Ebene schätzen. Zu diesem Zweck verwenden wir ein Area-Level-Modell, das Fay-Herriot-Modell, in Kombination mit Kovariablen aus Mobilfunkdaten. Da bei der Schätzung der Erwerbslosenquote das Pendlerverhalten der Erwerbsbevölkerung nicht einbezogen wird, verursacht dies ein verzerrtes Bild hinsichtlich der Verteilung der Erwerbslosenquote. Um dies zu berücksichtigen wird eine alternative Schätzung auf Basis der Arbeitsortes der Erwerbsbevölkerung für die Funktionalen Städtischen Gebiete in Nordrhein-Westfalen durchgeführt. Dadurch kann das Pendlerverhalten einbezogen und die Stadt-Land-Differenz ausgeglichen werden.

17 Robuste Schätzer für das Fay-Herriot-Modell

Tobias Schoch

University of Applied Sciences Northwestern Switzerland
tobias.schoch@fhnw.ch

SECTION: Methodology of Statistical Surveys

Robuste Methoden für die Small-Area-Schätzung von Mittel- und Totalwerten sind seit einiger Zeit bekannt und werden erfolgreich in der Praxis eingesetzt. Eine Vielzahl von „robustifizierten“ SAE-Schätzern ist aus ad-hoc-Überlegungen entstanden (was deren Tauglichkeit nicht schmälert). Für die Robustifizierung von Schätzern zum Fay-Herriot-Modell nehmen wir eine „theorienahe“ Perspektive ein, was zu neuen Einsichten führt.

Fay-Herriot (1979, J Amer Stat Assoc) haben das nach ihnen benannte Modell als Verallgemeinerung des James-Stein-Schätzers motiviert, wobei sie sich einen empirischen Bayes-Ansatz zunutze machten. Wir greifen diese Motivation des Problems auf und formulieren ein analoges robustes Bayes'sches Verfahren. Wählt man nun in der Bayes'schen Problemformulierung die ungünstigste Verteilung (eng. least favorable distribution) von Huber (1964, Ann Math Statist) als A-priori-Verteilung für die Lokationswerte der Small Areas, dann resultiert als Bayes-Schätzer [= Schätzer mit dem kleinsten Bayes-Risk] die Limited-Translation-Rule von Efron und Morris (1971, J Amer Stat Assoc). Im Kontext der frequentistischen Inferenzstatistik kann die Limited-translation-rule (LTR) nicht verwendet werden, weil sie als Bayes-Schätzer auf unbekanntem Parametern beruht. Diese unbekanntem Parameter können jedoch nach dem empirischen Bayes-Ansatz an der Randverteilung der abhängigen Variable geschätzt werden. Hierbei gilt es zu beachten (und dies wurde in der Literatur vernachlässigt), dass die Randverteilung unter der ungünstigsten A-priori-Verteilung nicht einer Normalverteilung entspricht, sondern durch die ungünstigste Verteilung nach Huber (1964) beschrieben wird. Bei den resultierenden Schätzern handelt es sich (und dies erstaunt nicht) um M-Schätzer, wie sie von Huber (1964) und weiteren Autoren vorgeschlagen wurden.

Unsere Diskussion zeigt auf, dass (1) die nach der M-Schätzermethodik geschätzte LTR bei (symmetrisch) kontaminierten Daten optimal ist und, dass (2) die LTR ein integraler Bestandteil der Schätzermethodik ist und nicht, wie es in der Literatur mitunter insinuiert wird, als „Zusatz“ verwendet werden kann. Wir illustrieren den Einsatz robuster Methoden an empirischen SAE-Beispielen und stellen ein R-Package vor.

18 Analysis of Poverty in a Village Level of Indonesia with Small Area Estimation: Case in Bangkalan District

Qonnita Silvia Makrufa; Ayesha Tantriana

Statistics Indonesia, Indonesia
ayeshatantriana@gmail.com

SECTION: Methodology of Statistical Surveys

Goals formulated in Sustainable Development Goals (SDGs) and national social policies indicate the need for analyzing poverty at local levels. Due to the data limitation, one of the ways to produce poverty indicator at local levels is by using Small Area Estimation (SAE) method. The estimation technique combines two data sources which are household sample surveys and the comprehensive coverage of a census. In measuring poverty, Statistics Indonesia uses the basic needs approach. The data sources are National Socio-economic Survey (SUSENAS) in 2016 and Population Census in 2010. The response variables in this study is poverty level at the level of sub-districts in Bangkalan District in East Java, Indonesia. To gain the poverty rate in sub-district or village level, this paper implements two approaches, Elbers, Elbers, Lanjou (EEL) method and Empirical Bayes (EB) method. These methods are compared to see which method is better in estimating the small area using SUSENAS and census data. The result shows that in the EB method, the villages that have the highest percentage of poor households are in Kanegarah village (70%) and the lowest is Mlajah village (9%). However, in the ELL method, village with the highest percentage of poor households is Kokop village (43.96%) and the lowest is Tanjung Jati village (1.42%). There are obvious differences in estimation results between the two methods, but not far enough. The Pearson correlation coefficient indicates that villages with high poverty rate in the EB estimation also have high poverty rate in the ELL estimation results. The result also shows that both EB and ELL method have a small bootstrap MSE value, indicating that both methods are good and reliable. To map the estimated poverty rate at sub-district level, this paper uses EB estimation as it has less MSE bootstrap and RRMSE score than EEL.

19 Bildungsmonitoring Berlin-Mitte – 2te Projektphase

Ulrike Rockmann

Senatsverwaltung für Inneres und Sport Berlin, Deutschland
ulrike.rockmann@uni-oldenburg.de

SECTION: Regional Statistics

– Elterliche Bildungsaspirationen der Einschulungskohorte 2019/2020 –

In der ersten Projektphase wurde aufgezeigt, dass regional wenig über die Bildungsaspirationen der Eltern von schulpflichtig werdenden Kindern bekannt ist, diese jedoch für den Bildungserfolg der Kinder eine wichtige Rolle spielen. Von Interesse ist, ob sich die Eltern für die Bildung ihrer Kinder interessieren, nach welchen Kriterien Bildungsinstitutionen beurteilt und wenn möglich ausgewählt werden. Um diese Aspekte zu beleuchten, wurde bei der Einschulungskohorte 2019/2020 in Berlin-Mitte ($n = 3.500$) im Zuge der Einschulungsuntersuchung eine Zusatzbefragung durchgeführt. Retrospektiv werden die Gründe für die Auswahl einer Kita beim Start des Besuches erfragt, aber auch, was den Eltern beim Kitabesuch kurz der Einschulung wichtig ist. Von Interesse ist auch, ob und wie sich die Eltern über die aufnehmende Grundschule informiert haben, welche Kenntnisse sie über das deutsche Bildungssystem haben und welche Lernumgebung das Kind in seiner Familie hat. Erste Befunde werden berichtet.

20 Sorting in an urban housing market – is there a response to demographic change?

Uwe Neumann; Lisa Taruttis

RWI - Leibniz-Institut für Wirtschaftsforschung, Germany
uwe.neumann@rwi-essen.de

SECTION: Regional Statistics

In urban areas, there is considerable neighbourhood-level variation in population characteristics. Using Dortmund as a case study we analyse whether and to what extent rents, housing prices and segregation dynamics corresponded with demographic ageing in urban neighbourhoods between 2007 and 2016. In a two-step analysis based on a hedonic pricing model and a discrete-choice model of housing location we find that in Dortmund so far there has been no slump of the housing market in neighbourhoods where the population ages more rapidly. Nevertheless, over the study period demographic segregation was on the rise and, according to a hedonic analysis, prices for apartments were higher in districts with a comparatively „younger“ population. In the course of further demographic change in Germany, which has come to a contemporary halt due to immigration, the response to ageing on urban housing markets in terms of location choice and prices may therefore become more evident. A large-scale urban regeneration project has revitalised the housing market of a declining Dortmund community during this decade. Since local ageing has not affected housing markets severely so far, it appears to be within the scope of urban policy to upgrade the attractiveness of ageing neighbourhoods as perceived by younger generations.

21 Die kleinräumige Perspektive: Disparitäten innerhalb von Städten

Teresa Lauerbach

Bundesinstitut für Bau-, Stadt- und Raumforschung, Deutschland
teresa.lauerbach@bbr.bund.de

SECTION: Regional Statistics

Der Begriff „Gleichwertige Lebensverhältnisse“ prägte 2018 viele politische Debatten. Insbesondere das Bundesinnenministerium unter Minister Horst Seehofer hat es sich mit der Gründung einer neuen Heimatabteilung für die aktuelle Legislaturperiode zur Aufgabe gemacht, die Lebensverhältnisse in Deutschlands Regionen anzugleichen und regionale Disparitäten abzubauen. Nach wie vor bestehen erhebliche Unterschiede in den regionalen Einkommens- und Beschäftigungsmöglichkeiten sowie bei der Sicherung der Mobilität und beim Zugang zu Angeboten der Grundversorgung und Daseinsvorsorge. Mit der Gründung der Kommission „Gleichwertige Lebensverhältnisse“ sollen nun Handlungsempfehlungen mit Blick auf unterschiedliche regionale Entwicklungen und den demografischen Wandel in Deutschland erarbeitet werden.

Hierüber bleibt hingegen bisweilen unberücksichtigt, dass auch auf kleinräumiger städtischer Ebene die Disparitäten nicht weniger, sondern sogar teilweise stärker ausgeprägt sind als auf regionaler Ebene. Dies ist insbesondere in Großstädten zu beobachten, in denen sich einzelne Stadtteile in der Zusammensetzung der in ihnen lebenden Bevölkerung stark voneinander unterscheiden. Vor allem das Problem der sozialräumlichen Polarisierung nimmt in vielen großen Städten weiter zu. Stadtteile mit Bewohnern mit überdurchschnittlich hohem Einkommen existieren direkt neben Stadtvierteln, deren Bevölkerung sich mehrheitlich aus Sozialleistungsempfängern zusammensetzt. Auch die hohen Zuwanderungsraten aus dem Ausland der letzten Jahre haben dazu beigetragen, dass kleinräumige Disparitäten weiter zugenommen haben.

Die Auswertungen zu innerstädtischen Disparitäten und Unterschieden zwischen Stadtteilen sollen auf Basis des Datenkatalogs „Innerstädtische Raumbewertung“ (IRB) erfolgen. Die kommunalen Statistikämter von 56 deutschen Großstädten liefern jährlich auf Basis einer Kooperationsvereinbarung Daten aus den Bereichen Bevölkerung, Wanderungen, Sozialtransfers, Beschäftigung und Arbeitslosigkeit sowie Bauen und Wohnen auf der kleinräumigen Ebene von Stadtteilen an das Bundesinstitut für Bau-, Stadt- und Raumforschung (BBSR). Die Daten reichen zurück bis zum Jahr 2002. Somit können zeitliche stadtübergreifende innerstädtische Entwicklungstrends auf kleinräumiger Ebene, die bei der Diskussion zur Gleichwertigkeit der Lebensverhältnisse nicht außer Acht gelassen werden sollten, beobachtet und analysiert werden.

22 Eine neue Symbiose von Stadt und Industrie? Vergleich regionaler Beschäftigungs- und Gründungsdynamiken in Deutschland

Ramona Voshage (1,3); Martin Gornig (2,3)

1: Amt für Statistik Berlin-Brandenburg; 2: Deutsches Institut für Wirtschaftsforschung (DIW Berlin); 3: Technische Universität Berlin
ramona.voshage@statistik-bbb.de

SECTION: Regional Statistics

Seit Mitte der 2000er Jahre ist mehr und mehr eine Renaissance der Großstädte zu beobachten. Im Fokus der Beobachtung steht dabei vor allem die Bevölkerungsentwicklung. Weniger weiß man hingegen über die wirtschaftliche Dynamik der Städte und insbesondere den Einfluss der Industrie, die einst die heutigen Großstädte geformt hatte.

In dem vorliegenden Beitrag wird gerade die industrielle Entwicklung in den Agglomerationen stärker in den Blick genommen. Ein Instrument für die Analyse der Arbeitsplatzdynamik ist die Betrachtung der Bruttoströme. Um diese Analyse durchzuführen, bedarf es einer Datengrundlage, die betrieblichen Einzeldaten in eine Gruppierung aller Betriebe in schließende, schrumpfende, neu gegründete und wachsende Betriebe ermöglicht. Das Ziel der empirischen Analyse der Arbeitsplatzdynamik besteht darin, die Nettoveränderungen im Beschäftigtenbestand im verarbeitenden Gewerbe in einzelne Komponenten zu zerlegen. Veränderungen im Bestand an beschäftigten Personen können das Ergebnis sehr verschiedener Prozesse sein: Der Abbau von Arbeitsplätzen kann durch Betriebsschließungen oder Betriebsschrumpfungen hervorgerufen werden. Arbeitsplätze entfallen, weil Betriebe ihre Tätigkeit einstellen oder ihren Personalbestand reduzieren. Zugleich entstehen immer wieder neue Arbeitsplätze. Zum einen, weil Betriebe neu gegründet werden und neue Arbeitsplätze schaffen oder zum anderen ihren Personalbestand erhöhen. Mit den verfügbaren und zusammengeführten Daten können Unterschiede zwischen Stadt und Land für das verarbeitende Gewerbe 2003 bis 2012 festgestellt werden.

Ein wesentlicher Frühindikator der Arbeitsplatzdynamik sind die Effekte, die von Gründungen ausgehen. Daher werden räumliche Analysen des Gründungsgeschehens in der Industrie für den Zeitraum 2012 bis 2016 vorgestellt. Sie geben Hinweise darauf, dass der aktuelle Erneuerungsprozess der Industrie insbesondere in den großen Städten stattfindet und dort die neue Attraktivität der Städte mitbegründet.

23 Implementing empirical results of panel models with lagged dependent variables and random intercepts into dynamic microsimulation

Petra Stein; Dawid Bekalarczyk; Eva Depenbrock

University of Duisburg-Essen, Germany
dawid.bekalarczyk@uni-due.de

SECTION: Microsimulations – Applications and Methods

Microsimulation models are commonly used to predict future developments in several societal areas (health, traffic, demographic transition etc.). The quality of such predictions strongly depends on the quality of the empirical input. In addition to the need of high-quality empirical starting data, the rules for updating this data during a microsimulation can also be derived from empirical data.

This brings up two main challenges. Firstly, developments on the individual level have to be estimated accurately. This can be done by dynamic panel models in which state dependence (effects of lagged dependent variables) is separated from time-invariant unobserved heterogeneity (modeled by a random intercept). However, it is difficult to obtain unbiased estimates in those models. For example, the inclusion of a lagged dependent variable (LDV) as an explanatory variable introduces an endogeneity bias, because the LDV is associated with the error term. Some approaches, which try to overcome these problems, will be presented. This includes models that use internal instruments (lags of already involved variables) (e.g. Arellano-Bond estimator using Generalized Method of Moments – Arellano, Bond 1991). Another method to solve endogeneity problems that will be discussed is a structural equation modeling approach (Williams, Allison, Moral-Benito 2018).

Secondly, it is challenging to transfer the results from these models into a microsimulation model when the starting data (simulation sample) does not coincide with the estimation sample, which is typically the case. The biggest difficulty arises from the fact that random intercepts are unknown for persons in the simulation sample. Richiardi (2014) developed a method to impute random intercepts by exploiting empirical information from the simulation sample (Rank Method). We will present this method and strategies to extend it to dynamic panel models. For several models, we have tested these implementation strategies with empirical and synthetic data.

24 Modelling the Unobserved: An Extension of Traditional Microsimulation

Sabine Zinn

SOEP DIW, Germany
szinn@diw.de

SECTION: Microsimulations – Applications and Methods

Dynamic microsimulation is a useful instrument to construct individual life-courses for whole populations. For its specification, typically transition models are used. Ideally, these transition models can be

estimated using statistical methods. However, for some transitions adequate data might not be available either because the corresponding processes are only partially observable (which is mostly the case for decision processes) or completely unobservable such as for future events. As a further impediment usual microsimulation is constrained to independent micro units making a straightforward modeling of network dynamics impossible. In this talk I describe a novel approach to enhance dynamic, purely data driven microsimulation to overcome both of these limitations. The main idea is to identify for transitions which cannot be determined by a statistical model proper sociological or psychological theories describing them well. Based on these theories, mathematical models are deduced such that transition rates or probabilities can be derived. This means that unprovable assumptions about the functional form of structural relationships have to be made. Thus, sensitivity analyses and validation are indispensable. To this end, I suggest using a meta-model of the enhanced microsimulation model such as a statistical emulator. I show the capabilities of the novel approach by investigating the effects of a hypothetical reform towards a more liberal education system in Bavaria on the propensity of primary school children to go to the Gymnasium (and not to another type of school) after the fourth grade.

25 Schätzung von Transitionswahrscheinlichkeiten für zeitdiskrete dynamische Mikrosimulationen unter struktureller regionaler Heterogenität

Joscha Krause; Jan Pablo Burgard; Ralf Münnich; Simon Schmaus

Universität Trier, Deutschland
krause@uni-trier.de

SECTION: Microsimulations – Applications and Methods

Mikrosimulationen sind ein mächtiges Instrument für die multivariate Analyse komplexer Systeme. Zunächst wird ein realitätsnahes synthetisches Abbild des zu erforschenden Systems auf der Mikroebene erstellt. Dann werden relevante Eigenschaften, welche den Ausgangszustand des Systems charakterisieren, szenarisch verändert. Die Effekte der Veränderungen werden jeweils über einen gegebenen Zeithorizont in die Zukunft projiziert. Der Vergleich der Szenarien gibt anschließend Aufschluss über wichtige Interdependenzen innerhalb des Systems. Im Folgenden werden zeitdiskrete dynamische Mikrosimulationen betrachtet. Sie zeichnen sich durch periodische stochastische Veränderungen der Mikroeinheiten im Zeitverlauf aus.

Um eine solche Mikrosimulation durchzuführen, müssen Transitionswahrscheinlichkeiten für alle Mikroeinheiten des Systems definiert werden. Sie quantifizieren die bedingte Wahrscheinlichkeit einer Zustandsänderung für eine spezifische Mikroeinheit gegeben seiner Charakteristika. Um authentische Simulationsergebnisse zu erhalten, müssen die Wahrscheinlichkeiten essenzielle Dynamiken des zu erforschenden Systems adäquat abbilden. Da Transitionswahrscheinlichkeiten meistens jedoch unbekannt sind, müssen sie aus geeigneten Stichprobendaten geschätzt werden. Dies erfolgt in der Regel auf Basis generalisierter linearer Modelle.

Die Schätzung von Transitionswahrscheinlichkeiten aus Stichprobendaten kann jedoch mit methodischen Problemen verbunden sein. Ist das zu erforschende System in regionale Teilsysteme gegliedert, welche sich hinsichtlich ihrer Transitionsdynamik strukturell unterscheiden, so muss die Heterogenität bei der Schätzung berücksichtigt werden. Dies ist in der Praxis jedoch häufig nicht möglich, da Stichproben-

daten gegebenenfalls nicht über hinreichend regionale Detailtiefe verfügen. Datenschutzbeschränkungen und begrenzte Erhebungsressourcen führen dazu, dass entweder keine regionale Verortung der Stichprobenelemente möglich ist, oder lokal nur wenige Beobachtungen vorliegen. In beiden Fällen kann die regionale Transitionsdynamik ohne methodische Anpassungen nicht adäquat abgebildet werden.

In diesem Beitrag diskutieren wir Erweiterungen generalisierter linearer Modelle zur Schätzung von Transitionswahrscheinlichkeiten für dynamische Mikrosimulationen. Die Erweiterungen sind dabei so konzipiert, dass sie strukturelle Unterschiede in der regionalen Transitionsdynamik erfassen, selbst wenn verfügbare Stichprobendaten nicht über genügend regionale Detailtiefe verfügen. Die Methoden werden zunächst theoretisch diskutiert, und anschließend in einer Simulationsstudie verglichen.

26 Synthesizing regression models to estimate micro-level data with multiple sources

Kristina M. Neufang; Ralf Münnich

Trier University, Germany
neufang@uni-trier.de

SECTION: Microsimulations – Applications and Methods

When setting up a microsimulation, it can be necessary to add research-relevant variables to the base dataset. Especially when registers are not available, survey data and other information have to be considered in order to obtain the required data. A possible way of adding a variable to the base dataset is the use of regression methods: A model can be estimated using auxiliary data in order to obtain fitted values for the variable of interest in the base dataset. When multiple sources of auxiliary data are available, each source can be used to estimate a specific regression model. Hence, it would be desirable to combine the single models in an overall model.

Addressing this problem, the current study applies methods originally used in the field of meta-analysis: The focus is on selected methods to synthesize coefficients in a linear regression context, i.e. a multivariate generalized least squares (GLS-) approach as well as a factored likelihood (FL-) method. To obtain informative findings by running a microsimulation detailed micro-level data is required. Hence, the current study evaluates both methods concerning their ability to increase the predictive power through the synthesis of (A) identical models (using GLS-Approach or FL-Method) and (B) nested models (using FL-Method) under different conditions (predictor structure, correlation among the variables, sample size) within a Monte-Carlo Simulation.

The findings indicate that both techniques increase the predictive power when synthesizing identical models. In case of synthesizing nested models, the performance of the FL-Method depends on the sample size of the full model. Since regression methods take an important part in the context of microsimulation, the use of synthesis methods may not only be advantageous in the described case. Additional fields of application are conceivable, like the estimation of transition probabilities with multiple sources.

27 Multivariate Conditional Transformation Models

Nadja Klein (2); Thomas Kneib (1); Torsten Hothorn (3)

1: Georg-August-Universität Göttingen, Germany; 2: Humboldt-Universität zu Berlin; 3: Universität Zürich
tkneib@uni-goettingen.de

SECTION: Computational Statistics, Data Analysis & Environmetrics

Regression models describing the joint distribution of multivariate response variables conditional on covariate information have become an important aspect of contemporary regression analysis. However, a limitation of such models is that they often rely on rather simplistic assumptions, e.g. a constant dependency structure that is not allowed to vary with the covariates. We propose a general framework for multivariate conditional transformation models that overcomes such limitations and describes the full joint distribution in simple, interpretable terms. Among the particular merits of the framework are that it can be embedded into likelihood-based inference and allows the dependence structure to vary with the covariates. In addition, the framework scales beyond bivariate response situations, which were the main focus of most earlier investigations. We illustrate the application of multivariate conditional transformation models in a trivariate analysis of childhood undernutrition and demonstrate empirically that even complex multivariate data-generating processes can be inferred from observations.

28 An Efficient, Unsupervised Learning Algorithm for Discovering Meaningful Clusters in Data

Hasan Hüseyin Oruc

FernUniversität in Hagen, Deutschland
hasan-hueseyin.oruc@fernuni-hagen.de

SECTION: Computational Statistics, Data Analysis & Environmetrics

In this paper we present a Bayesian Hierarchical Clustering algorithm (bhc), which can also be considered as performing approximate inference in a Dirichlet Process Mixture model. Our aim is to address drawbacks of traditional hierarchical methods, while being efficient and retaining desirable properties. More precisely, the bhc enables predictions about new data points by constructing a hierarchical representation of the data, and we can automatically discover interesting levels of the hierarchy. The algorithm is based on evaluating marginal likelihoods of a probabilistic model and merges clusters by using model-based criterion.

29 Adaptive LASSO estimation of the spatial weights matrix

Miryam Sarah Merk

European University Viadrina, Germany
merk@europa-uni.de

SECTION: Computational Statistics, Data Analysis & Environmetrics

Spatial econometric research is typically based on prior knowledge of the spatial weights matrix that characterizes cross-sectional interactions. For lattice data, spatial weights are commonly determined by contiguity, assumptions about nearest neighbors, as well as economic or social characteristics. These a priori specifications may lead to misspecifications of the model parameters, which are sensitive to the choice of the spatial weights. We therefore propose to select and estimate spatial dependence structures by using an adaptive Least Absolute Shrinkage and Selection Operator (LASSO). In the case of spatio-temporal models, the spatial dependencies of the process can be identified based on their observations over time. However, for purely spatial models the number of spatial links exceeds the sample size. The spatial weights are therefore estimated by cross-sectional resampling under the identifying assumption of sparsity.

30 Regional flood frequency analysis of possibly heterogeneous groups

Jona Lilienthal (1); Axel Bücher (2); Roland Fried (1)

1: Technische Universität Dortmund, Germany; 2: Heinrich-Heine-Universität Düsseldorf, Germany
lilienthal@statistik.tu-dortmund.de

SECTION: Computational Statistics, Data Analysis & Environmetrics

In the analysis of river flood events one is typically interested in estimating the recurrence time of rare flood events. A popular framework is the block maxima approach, i.e. only the maximum of each year is used and the distribution of this maximum is modelled by a generalized extreme value distribution. A common problem hereby is the shortage of data that leads to a volatile estimation of the shape of the distribution. An approach to improve this situation is regional flood frequency analysis, in which groups of similar stations are built and distribution parameters are estimated group-wise via L-moments using the Index Flood model.

The talk is about different aspects of regional flood frequency analysis. An important step is to verify the assumption of the Index Flood model. Known disadvantages of the commonly used homogeneity measure of Hosking and Wallis can be improved by the use of copula models and more robust trimmed L-moments. Often homogeneity of a set of stations is uncertain. In this case joint estimation can lead to less efficient quantile estimates. We propose penalized maximum-likelihood estimation, adding a regularization term to the likelihood function that is inspired by the Index Flood model. This allows for adjusting between local and regional calculation by a smoothing parameter. In simulation studies local and regional quantile estimations are compared to our proposed method using a cross-validated smoothing parameter.

31 Evaluating Effects of Tuition Fees: Lasso for the Case of Germany

Konstantin Görgen; Melanie Schienle

Karlsruhe Institute of Technology, Germany
melanie.schienle@kit.edu

SECTION: Empirical Economics

We study the effect of the introduction of university tuition fees on the enrollment behavior of students in Germany and determine a significant negative impact of fees on student enrollment. We do this by using an appropriate Lasso-technique that is crucial for identification of the magnitude and significance of the effect in this case with potentially many relevant controlling factors and only a short time frame where fees existed. Furthermore, we explicitly control for spatial cross-effects capturing the heterogeneity in the introduction scheme of fees that originated from the ability of each federal state („Bundesland“) to set its own educational policy. We show that a post-double selection strategy combined with stability selection identifies the true influencing variables, while a standard linear panel regression is not able to even determine the magnitude of the effect of tuition fees in this case.

32 Heterogeneity in Network Peer Effects

Livia Shkoza (1); Derya Uysal (2); Winfried Pohlmeier (1)

1: University of Konstanz, Germany; 2: University of Munich, Germany
livia.shkoza@uni-konstanz.de

SECTION: Empirical Economics

This paper studies peer effects in heterogeneous networks. In a composite model allowing for peer effects arising from both, local-aggregate and local-average information, we allow for further network heterogeneity, such that peer effects can differ across networks depending on exogenous factors. A novel Instrumental Variables – Minimum Distance estimator (IV-MDE) is proposed, which nests standard homogeneous peer effects model approaches. Our empirical application is based on network data of 90 school classes of secondary schools in Germany. Our findings indicate that ignoring network heterogeneity leads to misleading interpretations. We show that important effects discussed in the literature on school performance, such as class size and gender composition, operate indirectly through class specific peer effects.

33 Estimation of Latent Network Flow in Bike-Sharing Systems from Station Feeds

Marc Schneble; Göran Kauermann

Ludwig-Maximilians-Universität München
marc.schneble@stat.uni-muenchen.de

SECTION: Empirical Economics

Estimation of latent network flow is a common problem in statistical network analysis. Often, we know at least the margins of the network, i.e. in- and outdegrees. In our paper, we develop a generalized mixed regression model to estimate integer temporal network flows if only the differences of in- and outdegrees are known. Since the differences contain just few information on the network, we also make use of exogenous variables such as weather conditions, some of them modeled semiparametrically. Estimation can be performed via an iterative penalized maximum likelihood approach.

A simulation study is used to show the performance of the predictions. We apply our model to the Vienna Bike-Sharing network of the year 2014. For practical purposes it is crucial to predict when and at which station there is a lack of bikes or an excess of bikes. For this application, our model is well suited since the results show that station- and time-specific effects can be estimated well. Unsurprisingly, it is harder to perform estimation of route-specific effects since we do not have any information on routes traveled in the network.

34 Regression-based Network Reconstruction with Nodal and Dyadic Covariates and Random Effects

Michael Lebacher; Göran Kauermann

Ludwig-Maximilians Universität München, Institut für Statistik
michael.lebacher@stat.uni-muenchen.de

SECTION: Empirical Economics

Network (or matrix) reconstruction is a general problem which occurs if the margins of a matrix are given and the matrix entries need to be predicted. In this paper we show that the predictions obtained from the iterative proportional fitting procedure (IPFP) or equivalently maximum entropy (ME) can be obtained by restricted maximum likelihood estimation relying on augmented Lagrangian optimization. Based on the equivalence we extend the framework of network reconstruction towards regression by allowing for exogenous covariates and random heterogeneity effects. The proposed estimation approach is compared with different competing methods for network reconstruction and matrix estimation. Exemplary, we apply the approach to interbank lending data, provided by the Bank for International Settlement (BIS). This dataset provides full knowledge of the real network and is therefore suitable to evaluate the predictions of our approach. It is shown that the inclusion of exogenous information allows for superior predictions. Additionally, the approach allows to obtain prediction intervals via bootstrap that can be used to quantify the uncertainty attached to the predictions.

35 Neue Daten für die amtliche Statistik: Verknüpfen von Daten aus dem Internet mit Daten einer Erhebung bei Beherbergungsbetrieben in Hessen: Herausforderungen, Methoden und Ergebnisse

Tobias Gramlich

Statistik Hessen (Hessisches Statistisches Landesamt), Deutschland
tobias.gramlich@statistik.hessen.de

SECTION: Regional Statistics

Neben der Nutzung von Verwaltungsdatenregistern untersucht die amtliche Statistik in zunehmenden Maße die Eignung neuer digitaler Daten aus anderen Quellen für ihre Zwecke.

Für verschiedenen Bereiche der amtlichen Wirtschaftsstatistik lassen im Internet öffentlich zugängliche und durch Webscraping gewonnene Daten ein großes Potenzial erahnen. Diese Daten sind ohne Aufwand für Befragte und ohne Auskunftspflicht verfügbar und nutzbar.

Die Ziele von Webscraping sind in diesem Fall die Kontrolle oder Vervollständigung der Grundgesamtheit einer Erhebung oder auch die Anreicherung einer Erhebung um zusätzliche Information für Zwecke der Plausibilisierung oder Erhöhung des Analysepotenzials. Dadurch soll die Qualität der Erhebung insgesamt erhöht werden, Datengrundlagen vervollständigt, Ergebnisse schneller veröffentlicht oder auch auskunftspflichtige Betriebe in Zukunft weiter entlastet werden, wenn erfragte Information aus anderen Quellen in ausreichender Qualität verfügbar sind.

Nachdem die öffentlich zugänglichen Daten gewonnen wurden, müssen diese mit den Daten in der amtlichen Statistik verknüpft werden, damit das Potenzial der Daten genutzt werden kann.

Der Beitrag beschreibt die Verknüpfung der Daten aus der hessischen „Monatserhebung im Tourismus“ mit Daten, die über Webscraping von großen Hotelbuchungsportalen gewonnen wurden. Die Monatserhebung im Tourismus ist eine monatliche Vollerhebung mit Abschneidegrenze: d. h. monatlich müssen ggf. neue Beherbergungsbetriebe und ihre Bettenzahl recherchiert werden und ggf. in die Befragung aufgenommen werden. Hier soll die verbreitete Internetpräsenz von Beherbergungsbetrieben auf Buchungs- oder Bewertungsplattformen genutzt werden, um diese Recherchen weiter zu automatisieren und den Aufwand zu minimieren.

Für eine Verknüpfung stehen keine eindeutigen Identifikatoren zur Verfügung (z.B. Register- oder Steuernummern), sodass eine Verknüpfung über nichteindeutige und möglicherweise fehlerbehaftete Identifikatoren wie Unterkunftsname und -adresse stattfinden muss. Der Beitrag beschreibt Ausgangslage, Herausforderungen, Methode und Ergebnis der Verknüpfung am Beispiel ausgewählter Internetportale.

36 Daten aus Mobilfunknetzwerken für die amtliche Statistik: ein Pilotprojekt in der hessischen Tourismusstatistik

Tobias Gramlich

Statistik Hessen (Hessisches Statistisches Landesamt), Deutschland
tobias.gramlich@statistik.hessen.de

SECTION: Regional Statistics

Neben der Nutzung von Verwaltungsregistern untersucht die amtliche Statistik zunehmend den Einsatz Daten Dritter für ihre Zwecke. Anonyme Daten aus Mobilfunknetzwerken sind ein Beispiel solcher „neuer digitaler Daten“ für die amtliche Statistik.

Mobilfunkgeräte sind nahezu überall in Gebrauch. Durch die hohe Durchdringung des Alltags und der Gesellschaft in allen Bereichen könnten anonymisierte Daten aus Mobilfunknetzwerken ein hohes Potenzial für statistische Zwecke bieten, z.B. Zählungen zu bestimmten Zeitpunkten an bestimmten Orten, um daraus Bestandsgrößen abzuleiten oder als Grundlage für Mobilitätsanalysen.

Die amtliche Statistik, von den Städten bis zu supranationalen Organisationen, untersucht den Einsatz dieser prozessproduzierten Daten Dritter für ihre Zwecke.

Der Beitrag beschreibt ein Pilotprojekt in Hessen, das die Eignung vollständig anonymer Daten aus Mobilfunknetzwerken für Zwecke der amtlichen Tourismusstatistik untersucht. Die von den statistischen Landesämtern durchgeführte „Monatserhebung im Tourismus“ erhebt für alle Beherbergungsbetriebe oberhalb einer Abschneidegrenze die Anzahl der Gästeankünfte und -übernachtungen, ergänzt auch durch das Merkmal „Land des Wohnorts der Gäste“. Sie ist eine wichtige Konjunkturstatistik und Datengrundlage für Planung, Marketing, Investitionen und Monitoring für verschiedene Akteure. Das Pilotprojekt untersucht, ob anonyme Daten aus Mobilfunknetzwerken diese Statistik ergänzen und ihre Qualität weiter erhöhen können, z.B. durch eine höhere Schnelligkeit der Ergebnisse, höhere Genauigkeit durch Einbeziehung von Gästen und Übernachtungen in Betrieben unterhalb der Abschneidegrenze, bei gleichzeitig weniger Aufwand für die befragten Beherbergungsbetriebe.

Daneben bietet diese Datenquelle ggf. das Potenzial, auch bislang von der amtlichen Statistik nicht erfasste Ströme des Tages- oder Ausflugstourismus abzubilden.

Das Hessische Statistische Landesamt hat für dieses Pilotprojekt – unter Förderung aus Mitteln der Digitalstrategie des Landes Hessen – eine Kooperation mit einem Anbieter von absolut anonymen Daten aus Mobilfunknetzwerken zur Aufbereitung solcher Daten geschlossen.

Der Beitrag beschreibt das Pilotprojekt und zeigt erste Ergebnisse.

37 Identifikation von Unternehmen mit E-Commerce-Aktivitäten in der amtlichen Statistik durch Webscraping und Machine Learning

Normen Peters

Hessisches Statistisches Landesamt, Germany
npeters@statistik.hessen.de

SECTION: Regional Statistics

Mit einem im Hessischen Statistischen Landesamt bereits erfolgreich angewendeten Verfahren der automatisierten Informationsextraktion werden öffentlich zugängliche Daten von Webseiten und Internetpräsenzen von Unternehmen aus der Erhebung zur „Informations- und Kommunikationstechnologie in Unternehmen“ extrahiert und den jeweiligen Stammdaten zugeordnet.

In dem Beitrag wird die Suche nach den Webseiten der Unternehmen sowie die Identifikation, Verknüpfung und Analyse der extrahierten Daten beschrieben. Da sich die Existenz eines elektronischen Versandhandels und somit eine E-Commerce-Aktivität nicht in Form eines einzigen Datums von einer Internetseite extrahieren lässt, sondern ein Konstrukt aus beispielsweise einem vorhandenen Warenkorb, einer elektronischen Bezahlungsmöglichkeit, einem Shop und einer Log-in-Möglichkeit ist, müssen mehrere unterschiedliche Daten zusammengeführt werden, um die Existenz von E-Commerce-Aktivitäten zu bestimmen. Der Betrieb eines elektronischen Versandhandels (Onlineshop), als Hinweis auf E-Commerce-Aktivitäten auf einer Webseite, wird prädiktiv mit verschiedenen Verfahren des vollüberwachten maschinellen Lernens automatisiert nachgewiesen und die Ergebnisse werden evaluiert. Zur Evaluierung stehen sowohl die Befragungsergebnisse der Erhebung zur „Informations- und Kommunikationstechnologie in Unternehmen“ als auch die zur Erstellung des Trainingsdatensatzes verwendeten manuell recherchierten und geprüften Daten zur Verfügung.

Der Einsatz von Webscraping-Verfahren und maschinellem Lernen, zur Anreicherung der amtlichen Statistik und zur Unterstützung der Statistikerstellungsprozesse, ist ein wichtiger Baustein bei der fortwährenden Digitalisierung in der amtlichen Statistik. Diese Technologien und Verfahren werden im Hessischen Statistischen Landesamt seit etwa einem Jahr – insbesondere beim Ermitteln neuer hessischer Beherbergungsbetriebe und beim Schätzen des Bettenangebots – innerhalb der Tourismusstatistik überprüft

https://statistik.hessen.de/sites/statistik.hessen.de/files/Webscraping_von_Unternehmenswebseiten.pdf

https://statistik.hessen.de/sites/statistik.hessen.de/files/Aufsatz_Webscraping_Beherbergungsstatistik_04_18.pdf

Die Ergebnisse zeigen, dass Webscraping und das maschinelle Lernen das Potenzial haben, in der amtlichen Statistik eingesetzt werden zu können. Die Verfahren könnten u.a. dabei helfen, Auskunftspflichtige zu entlasten sowie die Produktvielfalt und deren Qualität zu erhöhen.

38 Fernerkundungsdaten in der amtlichen Statistik

Clara Maria Schartner

Destatis, Germany
clara.schartner@destatis.de

SECTION: Regional Statistics

Aus Fernerkundungsdaten lassen sich Merkmale ableiten, die für amtliche Statistiken genutzt werden können. Der Vortrag geht auf die Projekte im Institut für Forschung und Entwicklung in der Bundesstatistik ein und wirft auf einen Blick auf das Potential von Satellitendaten. Zurzeit laufen mehrere Projekte im Institut für Forschung und Entwicklung in der Bundesstatistik, die das Potential von Fernerkundungsdaten in verschiedenen Bereichen untersuchen. So werden zum Beispiel in einem Projekt mit ISTAT, CBS, HCSO, Universität Trier, Universität Pisa und Universität Southampton die Möglichkeit der Bestimmung von Nachhaltigkeitsindikatoren untersucht. In einem weiteren Projekt mit CBS, Statbel, IT.NRW und BISS wird die Erkennung von Solarpanels mithilfe von Fernerkundungsdaten untersucht. Des Weiteren wurde kürzlich ein Projekt abgeschlossen, das die technischen Möglichkeiten der Quantifizierungen von ökonomischen Aktivitäten unter Verwendung von Satellitendaten untersuchte. Mithilfe der Verschneidung von verschiedenen Datenquellen können Themengebiete umfassender analysiert werden. So wird im Rahmen des ESSnet Big Data II die urbane Lebensqualität durch die Verknüpfung von Zensus, Geo- und Fernerkundungsdaten untersucht. Durch das Zusammenführen von demographischen und geographischen Merkmalen von Stadtstrukturen können Stadtstrukturen viel umfassender betrachtet werden. Der Zensus 2011 ist auf 100 mal 100 Meter georeferenziert und beinhaltet Informationen über die demographischen Strukturen innerhalb dieser Gitterzellen wie zum Beispiel Familienstand, Alter und Geburtsland. Aus Fernerkundungsdaten kann die Landnutzung abgeleitet werden, so können z.B. Grünflächen von verbauter Fläche unterschieden werden. Außerdem beleuchten weitere Geodaten zusätzliche Facetten wie die Erreichbarkeit, die Abdeckung von Kinderbetreuungsstätten und kulturellen Einrichtungen. Durch die Zusammenführung dieser Datenquellen können die städtischen Strukturen multidimensional betrachtet werden.

39 Kontinuierliches Lernen in Einführungen in die Statistik durch intelligente Systeme unterstützen

Florian Berens; Sebastian Hobert

University of Goettingen, Germany
florian.berens@uni-goettingen.de

SECTION: Education and Training

Einführungen in die Statistik bestehen vielerorts aus Vorlesungen, kombiniert mit einem Tutoriums- oder Übungsbetrieb. Diese Präsenzveranstaltungen erfordern jedoch, dass Studierende hohes Engagement im Selbststudium zeigen, für das auch in den Prüfungs- und Studienordnungen ein großer Teil der Lernleistung vorgesehen ist.

Studierenden, insbesondere in der Studieneingangsphase, fällt es jedoch schwer, das eigene Lernen so zu organisieren, dass das Selbststudium zur regelmäßigen Unterstützung des Präsenzstudiums ausreicht. Das vorgestellte Projekt „Interactive Learning on Demand“ möchte Studierende in dieser Situation unterstützen.

Ansatzpunkt ist dabei die Nähe der Betreuung von Studierenden. In Vorlesungen ist Kommunikation zwischen Studierenden und Dozierenden zu individuellen Fragen schwer möglich. Tutorien bietet für individuelle Lernwege und Fragen zwar etwas mehr Raum, können aber ebenso wenig jeden Studierenden individuell begleiten. Ist in beiden Präsenzveranstaltungen noch direkter Kontakt vorhanden, ist der Zugang zu Hilfe in der Selbstlernphase noch schwieriger und im besten Falle sehr asynchron.

Das Projekt stellt für dieses Problem eine App zur Verfügung, die alle Lernschritte und Lernorte unterstützt. In Form eines Chats, der automatisch mittels Methoden der künstlichen Intelligenz antwortet, wird der Studierende jederzeit betreut und gefordert. Er wird durch Audience Response Systeme in Präsenzveranstaltungen und durch Quizze in der Selbstlernphase herausgefordert und durch die Möglichkeit der Push-Funktion zu regelmäßigem Lernen ermuntert. Elemente von Gamification sind an natürlichsprachliche positive Verstärkung angelehnt und verstärken die Lernmotivation.

Der Vortrag stellt sowohl die implementierte Software als auch deren Einbettung in das größere Lernkonzept vor und präsentiert erste Ergebnisse der Evaluation des ersten Probeseesters.

40 Messung statistischer Kompetenz am Ende eines Einführungskurses in die Statistik

Florian Berens

University of Goettingen, Germany
florian.berens@uni-goettingen.de

SECTION: Education and Training

Kernziel jeder statistischen Bildung ist es, statistische Kompetenz in einem umfassenden Sinne zu fördern. Dabei sind mindestens drei Gruppen zu unterscheiden: Für die Allgemeinbevölkerung sind statistische Kompetenzen relevant, die vor allem kritische Rezeption und Reflexion statistischer Information

ermöglichen. Für die Testung solcher Kompetenzen liegen bisher wenige geeignete Instrumente vor, Robert delMas und Kollegen liefern mit dem Statistics Reasoning Assesment aber eine gute Option.

Eine zweite Gruppe besteht aus professionellen Statistikern, also Personen, die Statistik zum Hauptberuf haben oder als Hauptfach studieren. Statistische Kompetenzen sind in dieser Gruppe von sehr diversem Umfang und daher schwer standardisiert zu erheben.

Dieser Beitrag möchte eine dritte Gruppe in den Fokus nehmen: Statistik wird in vielen/fast allen Wissenschaftszweigen angewandt und daher in vielen Studiengängen als Teil des Pflichtprogramms gelehrt. Diese Studierenden werden in die Grundlagen der Statistik eingeführt und nutzen Statistik danach in Anwendungsbezügen ihrer Fächer. Relevant für die Disziplin Statistik ist an dieser Stelle insbesondere die Frage, welche ihrer Elemente sie zu ihren Grundlagen zählt und welche statistischen Kompetenzen mit diesen Grundlagen ausgebildet werden sollen.

Für die USA konnte gezeigt werden, dass die statistischen Grundkurse des Landes zu etwa 80% dieselben Inhalte abdecken. Es scheint damit trotz fehlender Vorgaben von oben eine hohe Standardisierung erreicht zu sein, die aber bisher nicht durch gemeinsame Standards ihrer Testung begleitet ist.

Der Vortrag stellt in diesem Kontext einen Entwurf zur Diskussion, der statistische Kompetenz am Ende eines Einführungskurses misst. Abgedeckt werden darin konzeptuelle, kalkulatorische wie interpretative Aspekte statistischer Kompetenz. Der Test schließt damit an die Debatten zu Statistical Reasoning, Statistical Thinking und Statistical Decision Making an und bricht diese Ansätze konkret herunter auf die drei Inhaltsbereiche univariate deskriptive Statistik, Grundlagen der Inferenz und Einfache bivariate Statistik.

41 Structured Eurobonds: Pricing the political cost of joint liability

Christian Bauer; Marc Adolph

Universität Trier, Germany
bauer@uni-trier.de

SECTION: Macroeconomics & Finance

In this paper we introduce an analytical tool to analyze the effects of an ABS-approach to implement structured Eurobonds. Structured Eurobonds could overcome the huge and emotionally empowered political obstacle of joint liability as they allow to choose a suitable degree of jointness. In addition, the interest gains created by the structure are distributed to participating countries, for AAA countries to compensate for the risks of participation and for the others as a financial benefit.

We show that the approach is stable over time, has an incentive compatible micro structure, meets the stylized facts and increases the volume of AAA bonds in the Eurozone in total.

In addition, we address the discussion about joint liabilities. Interest savings reach a maximum of 0.5 percentage points depending on the degree of joint liability and other global parameters like the assumed recovery rate and global riskless interest rate. Each member participates in the joint liability only with a certain share of its individual liability which lies between 10% and 17% in the optimum.

Using these results, we can price the political cost of the degree of joint liability. In all scenarios, joint liability shows a decreasing rate of return from positive to negative.

In the sensitivity analysis, we analyze different scenarios: a participation of all countries vs. the drop out of single countries or a small subset as the PIIGS only. Marginal diversification and tranching effects are decreasing in the number of countries. We show that countries with high interest load are most profitable for the community.

Summarizing, structured Eurobonds are a stable, very beneficial and political feasible tool for a European fiscal system. The implementation with a supranational organization as the ESM to ensure low issuer risk is especially important to create financial benefits if the global interest rate is low.

42 Towards integrating Fintech into statistical classification systems – a process oriented approach

Ulf von Kalckreuth; Norman Wilson

Deutsche Bundesbank, Germany
ulf.von-kalckreuth@bundesbank.de

SECTION: Macroeconomics & Finance

In the course of upcoming revisions of international classification standards, this paper collects ideas on how to adapt classification systems such as the CPC for products and the ISIC for activities to the advent of „Fintech“ firms and their position in the financial system. The theoretical argument is complemented with an empirical study of the distribution of the official statistical sector classification of Fintech firms in Germany according to the standards valid today.

There are two key results: First, it will not be possible to take the definition of „Fintech“ currently used in the literature for statistical classification. The reason is that for the definitions of „Fintech“ currently in use, the two concepts „innovative“ and „technology“ are absolutely essential. These concepts are not time-invariant. A firm that is „Fintech“ today may not be „Fintech“ ten years from now if it continues to provide the same services. Second, in order to enable statistics to register and map financial activity in the economy, a useful way is to identify the key processes necessary to produce financial services and to classify those activities and products as „financial“, no matter whether they are provided within a traditional financial institution or in IT companies specialised on one link of the financial value chain. This will make statistical measurement immune against outsourcing and specialisation within the financial industry and at the same time also open for rapid technological progress in the future. Ultimately, to cope with BigTechs, statistical reporting obligations on financial issues should not depend on the industry classification of a firm.

43 Statistical Governance and FDI in Emerging Economies

Ulf von Kalckreuth

Deutsche Bundesbank, Germany
ulf.von-kalckreuth@bundesbank.de

SECTION: Macroeconomics & Finance

This paper investigates the role of official statistics for alleviating financing constraints with regard to funding investment projects in emerging economies. Official Statistics has an important dual role: it adds to the information set of investors regarding the general state of the economy and it is an important commitment and signalling device as to future good governance. A statistical system according to international standards is costly to set up, but facilitates monitoring the government policy. This renders credible the government's pledge to keep up transparency and good governance in the future.

Empirically, the paper investigates, for a sample of 98 emerging and developing countries, the relationship between the adoption of the IMF General Data Dissemination Standard (GDSD) for statistical data production on the one hand and the net incurrence of foreign direct investment liabilities on the other. Considering the conditional distributions of country-years, direct investment is sizeably higher under GDSD. Controlling also for time and country effects, using fixed effects and quantile panel regression, the relationship ceases to be uniformly positive. Heterogeneity matters: There is a large and significant difference between poorer countries and richer countries. It turns out that the relationship between the adoption of GDSD and net incurrence of FDI liabilities is actually negative for richer countries and outside Sub-Saharan Africa. For richer countries, the relevant alternative might have been the more demanding SDDS standard, turning the adoption of GDSD into an unfavourable signal.

Quantile regression is carried out using the fixed effects quantile panel estimator of Canay (2011)

44 Social contagion processes on networks: which random graph model fits best?

Tyll Krüger

NA
NA

SECTION: Microsimulations – Applications and Methods

Real social networks are instances of a random graph generating process whose details are usually only poorly known. It is a challenging statistical and modelling problem to choose the right random graph space for a given observed network. Besides the requirement that the random graph family should reflect the typical patterns of a given network one is often interested in modelling social contagion processes taking place on those networks and predict their dynamics. Typical examples are: the spread of information and knowledge, the diffusion of opinions or the spread of social norms. We present some recent classes of random graph spaces and discuss to which extend in a statistical reliable way their defining features can be estimated from large network samples. Furthermore we present some theoretical and simulation results for epidemic contagion processes, discuss their relevance in the context of microsimulations and give applications of the theory for the prediction of Twitter cascades.

45 Distributed Survival Analysis under Disclosure Constraints

Federico Bonfiglio (1,2); Martin vSchumacher (1,2); Harald Binder (1,2)

1: Albert-Ludwigs-Universität Freiburg, University Medical Center Freiburg, Germany; 2: Freiburg Center for Data Analysis and Modeling, Germany
bono@imbi.uni-freiburg.de

SECTION: Microsimulations – Applications and Methods

We consider pooling of original Individual Participant Data (IPD) between different locations/servers under disclosure constraints. That is, locations cannot share original IPD but only anonymous IPD proxies or summaries, that are later pooled. This procedure is known as distributed computing (also implemented in the DataSHIELD infrastructure) and obtaining original IPD inferences while protecting privacy can be difficult here. For instance, to compute IPD Cox (frailty) regression or Nelson-Aalen/Breslow estimates is currently impossible here.

We present a micro-simulation method to readily approximate such survival inferences in a privacy-constrained distributed network like DataSHIELD.

The idea is to first generate a private version of the original IPD and then to compute IPD inferences on it. Private IPD reconstructions are generated via a copula inversion technique that uses empirical IPD marginal moments and correlation matrix, disclosed by each location, as input data only. We show with practical examples that we can well recover fixed and latent effect estimates of an IPD Cox regression, as well as Nelson-Aalen estimates, from such private IPD reconstructions whose broader utility we comment upon.

46 Random Fields evolving temporally over spheres

Emilio Porcu

Newcastle University, UK
emilio.porcu@newcastle.ac.uk

SECTION: Young-Academics Mini-Symposium: Philipp Otto: Models for Spatial and Spatiotemporal Data

This talk introduces random fields defined over spheres and evolving over time. Special attention on statistical modeling techniques for space-time processes, where space is the sphere representing our planet, is given. In particular, we make a distinction between (a) second order-based approaches and (b) practical approaches to modeling temporally evolving global processes. The former approaches are based on the specification of a class of space-time covariance functions, with space being the two-dimensional sphere. The latter are based on explicit description of the dynamics of the space-time process, that is, by specifying its evolution as a function of its past history with added spatially dependent noise.

Special emphasis is put on approach (a), albeit practical approaches will be discussed as well.

We present a case study focused on the analysis of air pollution from the 2015 wildfires in Equatorial Asia, an event that was classified as the year's worst environmental disaster. The paper finishes with a list of the main theoretical and applied research problems in the area, where we expect the statistical community to engage over the next decade.

47 Dynamic Spatial Autoregressive Models with Time-varying Spatial Weighting Matrices

Anna Gloria Billé(1); Francisco Blasques (2); Leopoldo Catania (3)

1: Free University of Bolzano-Bozen, Italy; 2: VU University Amsterdam; 3: Aarhus University
annagloria.bille@unibz.it

SECTION: Young-Academics Mini-Symposium: Philipp Otto: Models for Spatial and Spatiotemporal Data

We propose a new spatio-temporal model with time-varying spatial weighting matrices. We allow for a general parameterization of the spatial matrix, such as: (i) a function of the inverse distances among pairs of units in space to the power of an unknown time-varying distance decay parameter, and (ii) a negative exponential function of the time-varying parameter as in (i). The filtering procedure of the time-varying unknown parameters is performed using the information contained in the score of the conditional distribution of the observables. We provide conditions for the stationarity and ergodicity of the filtered sequence of the spatial matrices as well as for the consistency and asymptotic normality of the maximum likelihood estimator (MLE).

An extensive Monte Carlo simulation study to investigate the finite sample properties of the MLE is reported. Finally, two empirical applications are reported. We first analyze the association between eight European countries' perceived risk, finding that the economically strong countries have their perceived

risk increased due to their spatial connection with the economically weaker countries. Secondly, we investigate the evolution of the spatial connection between the house prices in different areas of the UK, identifying periods when the usually adopted sparse weighting matrix is not sufficient to describe the underlying spatial process.

48 Interpolation of Weather Conditions for a Flight Corridor

Gong Chen; Ostap Okhrin

Technische Universität Dresden, Germany
gong.chen1@tu-dresden.de

SECTION: Young-Academics Mini-Symposium: Philipp Otto: Models for Spatial and Spatiotemporal Data

The main issue comes from the air logistic where an aircraft needs to optimize trajectories to reduce emission and to avoid dangerous weather conditions. This work provides an interpolation of weather variables like temperature, wind speed, wind direction and pressure. Data from the Global Forecast System (GFS) is unfortunately too sparse for dynamic trajectory optimization thus cannot be used directly. Weather variables are categorized as scalar and vector for the reason that wind has two dimensions, wind speed and wind direction. For each variable, interpolation methods, including Inverse Distance Weight, Kriging and polynomial method, are compared using a cross validation procedure. Universal Kriging is applied rather than Ordinary Kriging because weather conditions vary geographically.

49 How should we filter economic time series?

Yves Schüler

Deutsche Bundesbank, Germany
yves.schueler@bundesbank.de

SECTION: Empirical Economics

Hamilton (2017) criticises the HP filter because of three drawbacks (i. spurious cycles, ii. end-of-sample bias, iii. ad hoc assumptions regarding the smoothing parameter) which, moreover, apply to other popular time series filters as well. As an alternative filter, he proposes a regression filter. I demonstrate that Hamilton's regression filter is partially subject to the same drawbacks (i. and iii.). Furthermore, I illustrate that Hamilton's regression filter does not fulfil established criteria for the desired properties of filters. For instance, in contrast to the HP filter, Hamilton's filter does not approximate the ideal band pass filter, does not have constant cyclical properties across time series, and induces phase shifts. I discuss two refinements to bridge Hamilton's and the established view. For extracting business or financial cycles, I find that a one-sided HP filter and a 1-quarter regression filter can best reconcile both perspectives.

50 Oracle Efficient Estimation of Structural Breaks in Cointegrating Regressions

Karsten Schweikert

University of Hohenheim, Germany
karsten.schweikert@uni-hohenheim.de

SECTION: Empirical Economics

In this paper, we propose an adaptive group lasso procedure to efficiently estimate structural breaks in cointegrating regressions. It is well-known that the group lasso estimator is not simultaneously estimation consistent and model selection consistent in structural break settings. Hence, we use a first step group lasso estimation of a diverging number of breakpoint candidates to produce weights for a second adaptive group lasso estimation. We prove that parameter changes are estimated consistently by group lasso if it is tuned correctly and show that the number of estimated breaks is greater than the true number but still sufficiently close to it. Then, we use these results and prove that the adaptive group lasso has oracle properties if weights are obtained from our first step estimation and the tuning parameter satisfies some further restrictions. Simulation results show that the proposed estimator delivers the expected results. An economic application to the long-run US money demand function demonstrates the practical importance of this methodology.

51 „Quo vadis, Statistik? – Eine GIS-gestützte Analyse der niedersächsischen Beherbergungswirtschaft“

Julian Schulz; Franziska Große

Landesamt für Statistik Niedersachsen, Deutschland
julian.schulz@statistik.niedersachsen.de
franziska.grosse@statistik.niedersachsen.de

SECTION: Regional Statistics

Abstract zu Auswertungen auf Basis georeferenzierter Daten
„Quo vadis, Statistik? – Eine GIS-gestützte Analyse der niedersächsischen Beherbergungswirtschaft“

Niedersachsen ist für inländische und ausländische Gäste als Reiseziel sehr beliebt. Jedoch ist die Tourismuslandschaft in einem Flächenland wie Niedersachsen regional sehr heterogen. Auf der einen Seite die klassischen Tourismusgebiete wie Nordseeküste, Lüneburger Heide und der Harz. Auf der anderen Seite Städte wie Hannover, Braunschweig oder Göttingen, die vom Städte- und/oder Messtourismus geprägt sind. Und dazwischen Landstriche, in denen wenige Reisende übernachten.

Der vorliegende Beitrag untersucht, welche Regionen vom Tourismus besonders stark geprägt sind und wie sich diese Verteilung GIS-gestützt analysieren lässt. Die Betrachtung dieser regionalen Disparitäten erfolgt anhand von georeferenzierten Daten der monatlichen Beherbergungsstatistik. Im Rahmen der Analyse werden folgende Fragestellungen untersucht: Wo liegen die Hotspots mit dem höchsten/niedrigsten Verhältnis von Bettenzahl pro Kopf und zeigen diese damit die regionale Bedeutung bzw. Abhängigkeit dieser Region vom Tourismus im Vergleich zur wirtschaftlichen Gesamtschau? Wie ist die Kerndichte der sozialversicherungspflichtig Beschäftigten (SVB) im Beherbergungsgewerbe relativ zu der Gesamtzahl der SVB in Niedersachsen verteilt? Das Ziel ist also eine Herangehensweise zu finden, wie durch kartographische Instrumente ein Mehrwert aus den bereits vorhandenen und in Tabellenform publizierten Maßzahlen geschaffen werden kann, und dies stets unter Berücksichtigung der statistischen Geheimhaltung.

Außerdem stellt sich die Frage, welche Auswirkungen der Tourismus auf weitere Wirtschaftsbereiche in den Regionen hat. Lässt sich in Tourismusregionen ein auffällig höherer Beschäftigtenanteil im Gastgewerbe als in anderen Regionen nachweisen? Welche weiteren Wirtschaftszweige sind über- oder unterrepräsentiert? Wie können diese Muster mit den Daten der amtlichen Statistik kartographisch analysiert werden? Die Ergebnisse können dazu beitragen, bei der regionalen Wirtschafts- und Tourismusförderung zielgerichteter vorzugehen, indem durch die Verwendung von Gitterzellen gezielt auch solche räumlichen Verteilungsmuster identifiziert werden, die sich nicht an den Gemeindegrenzen orientieren.

52 Neue Pendleranalysen mit Mobilfunkdaten

Günther Bachmann

Amt für Wirtschaft und Stadtentwicklung, Deutschland
guenther.bachmann@darmstadt.de

SECTION: Regional Statistics

Die Wissenschaftsstadt Darmstadt hat – ähnlich wie andere deutsche Großstädte – erhebliche Problemlagen durch eine hohe Feinstaub- und Stickoxidbelastung in der Innenstadt sowie große Lärmlasten an den Hauptverkehrsstraßen, hauptsächlich durch Pendler in der und in die Stadt.

Nach dem Gewinn des 1. Platzes beim Bitkom-Wettbewerb „Digitale Stadt“ eröffnen sich derzeit gänzlich neue Wege zur Optimierung der Mobilitätsketten mit Projekten der Smart City Darmstadt. Am Beispiel der Auswertung hochdifferenziert aufgelöster räumlicher und zeitlicher Mobilfunkdaten ist die große Chance gegeben, durch modernste Verfahren die Mobilitätswege in und nach der Stadt zu optimieren und über neue Verkehrslösungen nachzudenken. Mittels neuartiger Algorithmen, die im Vortrag vorgestellt werden, lassen sich die Pendlerbewegungen innerhalb der Stadt und von außen wesentlich besser abbilden und ergänzen damit die vorhandenen amtlichen Statistikdaten um wichtige neue Erkenntnisse. Auch die Pendlerbewegungen von und zu Arbeitsstätten sind damit erstmals in hoher Detailliertheit und Schärfe abbildbar und bieten die große Chance für eine bessere umweltgerechte und sozialverträgliche Verkehrsplanung in der Wissenschaftsstadt Darmstadt.

53 Seminar „Statistical Literacy“ – Statistikwissen nutzbar machen

Kevin Stabenow; Friederike Schmal

WWU Münster, Deutschland
kevin.stabenow@wiwi.uni-muenster.de

SECTION: Education and Training

Als Pflichtfach des ersten Studienjahres verkommt Statistik für viele Studierende zu einer ungeliebten Pflichtveranstaltung fern von jeglicher Anwendung und kritischer Auseinandersetzung mit dem Gelernten. Vielmehr werden intensiv Formeln für Kovarianzen, Rangkorrelationskoeffizienten und Teststatistiken auswendig gelernt und kochrezeptartig Rechenwege aus Tutorien in Klausuren reproduziert. Insbesondere für die von uns unterrichteten Wirtschaftswissenschaftler_innen erscheint daraus eher ein Kosten-Nutzen-Kalkül zu entstehen in Richtung „Wie kann ich bei einem geringen Lernaufwand ein möglichst gutes Ergebnis erzielen?“ als dass ein gutes Gespür für Statistiken, Statistical Literacy, erwächst. Auch in Vertiefungsmodulen aus dem Wahlbereich wird viel Lehrzeit auf das Wiederholen (Neubeibringen) von mathematischem und statistischem Grundlagenwissen verwendet und das Augenmerk fällt nur wenig auf die Anwendung von jenen Methoden im „echten“ Leben – Wissenschaft, Journalismus und Werbung.

Das Institut für Ökonometrie und Wirtschaftsstatistik der Wirtschaftswissenschaftlichen Fakultät der WWU Münster versucht diesem Problem in der Lehre entgegenzuwirken durch ein neues Seminarangebot „Statistical Literacy – Statistische Fehlentscheidungen und Konsequenzen“. Ziel des Seminars ist es, eine wissenschaftliche Veröffentlichung oder die Medienaufarbeitung einer Veröffentlichung hinsichtlich der darin genutzten statistischen Methoden und Interpretationen durchzuarbeiten. Besonderer Schwerpunkt soll dabei auf mögliche Fehler in der Methodenwahl, Ausführung und Interpretation gelegt werden. Dabei kann es sich beispielsweise um die Verwechslung von Kausalität und Korrelation, fehlende Signifikanzniveaus, Bias bei Antworten oder Confounding handeln.

Während für Bachelor-Studierende der Fokus auf der Analyse der Veröffentlichung und der Kritik an dieser liegen soll, wird von Master-Studierenden eine empirische Arbeit mit Ziel der Reproduktion und Verbesserung der Veröffentlichung erwartet. Als Lehrende erhoffen wir uns für die Studierenden inhaltlich ein besseres (intuitiveres) Verständnis von Statistiken und Medienreaktionen, einen erleichterten Einstieg in das Angehen von (größeren) wissenschaftlichen Arbeiten wie der Bachelor- oder Masterarbeit und einen Ausbau der Softwareskills (R und Stata).

Ziel des Vortrags ist es unsere Erfahrungen mit der Durchführung des Seminars zu teilen und eine Diskussion über das Unterrichtsformat anzuregen.

54 Lernzentrierte Übungssystematik und statistisches Monitoring

Toni Stocker

Universität Mannheim, Germany
stocker@uni-mannheim.de

SECTION: Education and Training

Das in vielen Fächern wichtige Lehrveranstaltungsformat der vorlesungsbegleitenden Übung wird häufig nicht in der Weise praktiziert wie es eigentlich wünschenswert wäre. So werden viele Übungen nach wie vor in einer Art Frontalstil abgehalten, bei dem in erster Linie nur Lösungen von Aufgaben präsentiert werden. Solche „Frontalübungen“ bergen jedoch einen hohen Grad an Ineffizienz, da die Studierenden eine überwiegend nur passive Rolle einnehmen und den individuellen Verständnisproblemen kaum begegnet werden kann. Als Alternative dazu wird hier eine Übungssystematik vorgestellt, im Rahmen derer die Studierenden ihren Lernprozess aktiv selbst bestimmen und begleitend dabei unterstützt werden. Der Ansatz besteht im Kern darin, die Heterogenität unter den Studierenden durch Bildung homogener Kleingruppen ein Stück weit aufzulösen, um auf diese Weise unterschiedliche Vorbereitungen, Verständnisprobleme und Lerngeschwindigkeiten individueller und effizienter berücksichtigen zu können. Dieser Ansatz wurde in Grundlagenveranstaltungen zum Fach Statistik über einen Zeitraum von nunmehr 10 Jahren zu einem in der Praxis sehr gut funktionierenden Gesamtkonzept ausgebaut, das auf allgemein hohe Akzeptanz unter den Studierenden stößt und zu messbar höheren Lernaktivitäten und Lernerfolgen führt als im Frontalsystem.

Im Rahmen dieses Lehr-Lern-Konzepts fallen regelmäßig Daten über den Vorbereitungsgrad und die Lernaktivität der Studierenden auf Individual- und Gruppenebene an, die systematisch erfasst werden. Anhand von Beispielen wird aufgezeigt, wie sich diese Daten dann in vielfältiger Weise im Rahmen eines statistischen Monitorings zum Zwecke der Deskription, Steuerung, Prognose und Evaluation nutzen lassen.

55 No-Screen-Policy in der Präsenzlehre – verbunden mit eLearning-Angeboten zum selbstgesteuerten Lernen

Christine Buchholz

Hochschule Bonn-Rhein-Sieg, Deutschland
christine.Buchholz@h-brs.de

SECTION: Education and Training

Die Digitalisierung ist im Hochschulalltag angekommen, für viele Studierende ist es selbstverständlich, auch in der Präsenzlehre online zu sein.

Einerseits können digitale Tools in der Präsenzlehre sinnvoll zur Aktivierung der Studierenden genutzt werden, andererseits besitzen mobile Endgeräten ein enormes Ablenkungspotenzial. Dies betrifft in erster Linie den Nutzer selbst, aber darüber hinaus auch Mitstudierende und Dozenten. Fraglich ist, welchen Einfluss das Multitasking der Nutzer auf die Lernatmosphäre und den Lernerfolg im Auditorium hat.

Dozenten stören sich mitunter an der Nutzung mobiler Endgeräte, aber selten werden klare Regeln hinsichtlich deren Nutzung für die Präsenzlehre aufgestellt.

Dieser Beitrag thematisiert die Beweggründe, eine No-Screen-Policy für die klassische Vorlesung aufzustellen und stellt wissenschaftliche Studienergebnisse sowie Erfahrungen in der praktischen Implementierung vor.

Darüber hinaus werden didaktische Elemente zur Aktivierung der Studierenden präsentiert und erläutert, wie dem Wunsch nach digitalem, selbstverantwortlichen Lernen mittels e-Learning-Angeboten Rechnung getragen wird.

56 Learning to understand and to do Statistics step-wise by R shiny, learnr, mosaic and markdown

Karsten Lübke; Matthias Gehrke

FOM, Germany
karsten.luebke@fom.de

SECTION: Education and Training

There are calls for an amplification in statistical education towards Data Literacy. Driven by digitalization today, there is more and different data available. Also technology has never provided as many opportunities for analyses as today. With this expanding data and data analysis universe we should rethink the curriculum as well as the technologies to best prepare our students to be data literate citizens in the 21st century. As Wild et al. (2017) put it: „With the rapid, ongoing expansions in the world of data, we need to devise ways of getting more students much further, much faster.“

In this talk we will present our building blocks towards data literacy education. For this we constructed an integrated framework from interactive apps (for understanding statistics) to reproducible analysis (for doing statistics), based on the key concepts of modeling and simulation based inference.

57 Drivers of individual performance of professional biathletes

Christoph Herrmann; Thomas Kirschstein; David Lukas

Martin-Luther-Universität Halle-Wittenberg, Germany
christoph.herrmann@wiwi.uni-halle.de

SECTION: Computational Statistics and Data Analysis: Applied Statistics

Sports analytics has professionalized drastically in recent years. Particularly in prominent sports like soccer and American football, professional data analysis is a cornerstone of training and sports management. In biathlon, however, there are only a few articles analyzing individual performance of biathletes. However, biathlon is a highly versatile sport consisting of Nordic skiing and shooting tasks. Thus, athletes are confronted with high physical as well as psychological demands. On the one hand, they have to ski as fast as possible and, moreover, they need also a well trained endurance. On the other hand, high precision and concentration including a „steady hand“ are required at the shooting tasks (prone and standing). Physical exhaustion and concentration are antagonists. This versatility motivates to take a closer look on the main attributes of individual performance in biathlon.

We use the results of all competitions in the disciplines „sprint“ and „individual“ from five International Biathlon Union world cup seasons to analyze individual performance of biathletes. Based on the empirical data we derive performance indicators for skiing and shooting which are analyzed for their predictive accuracy for different types of races. Finally, we study the effects of course characteristics as well as weather and snow conditions on individual performance.

58 PISA Performance of Natives and Immigrants: Selection versus Efficiency

Andreas Behr; Gerald Fugger

Universität-Duisburg Essen, Deutschland
gerald.fugger@uni-due.de

SECTION: Computational Statistics and Data Analysis: Applied Statistics

In most countries, immigrant and native students perform differently in PISA due to two main reasons: different immigration regimes and differences in their home-country educational systems. While there is sophisticated literature on the reasons for these performance gaps, it is barely considered in the educational efficiency research. Our approach distinguishes between selection effects caused by immigration policies, and the efficiency of educational systems in integrating immigrant students, given their socio-economic background. Accordingly, we split our sample, which consists of 153,374 students in 20 countries, calculate various different efficient frontiers, and ultimately decompose and interpret the resulting efficiency values. We find large differences in educational system efficiency, when we compensate for negative selection effects caused by immigration regimes.

59 Is there a hot hand effect in professional darts? An analysis using state-space models

Marius Ötting (1); Roland Langrock (1); Christian Deutscher (1); Vianey Leos Barajas (2)

1: Bielefeld University, Deutschland; 2: Iowa State University, USA
roland.langrock@uni-bielefeld.de

SECTION: Computational Statistics and Data Analysis: Applied Statistics

In sports, the concept of the „hot hand“ refers to the idea that athletes may enter a state in which they experience exceptional success. For example, in basketball, players are commonly referred to as being „in the zone“ or „on fire“ when they hit several shots in a row. Although players, managers and fans strongly believe in its existence, the hot hand has been the subject of intense scientific debate over more than three decades, with some researchers attributing the alleged effect to a cognitive illusion.

Here we investigate the hot hand hypothesis in professional darts as a near-ideal setting with minimal to no interaction between players. Considering almost one year of tournament data, corresponding to 167,492 dart throws in total, state-space models are used to investigate serial dependence in throwing performances. In these models, a latent state process serves as a proxy for a player’s underlying ability, and autoregressive processes are used to model how this process evolves over time. The results show a strong but short-lived serial dependence in the latent state process, thus providing evidence for the existence of the hot hand.

60 Rent Control and Rental Prices: High Expectations, High Effectiveness?

Philipp Breidenbach; Lea Eilers; Jan Fries

1: RWI – Leibniz-Institut für Wirtschaftsforschung, Deutschland; 2: SVR Wirtschaft
breidenbach@rwi-essen.de

SECTION: Poster

This paper evaluates the rent control policy implemented in Germany in 2015. Like many countries around the world, German cities and metropolitan areas have experienced a strong increase in rental prices during the last decade. In response, the politicians aimed to dampen the rise in rental prices by limiting the landlords' freedom to increase rents for new contracts. To that end, the rent control was introduced. To evaluate the effectiveness of the rent control with respect to rental prices, we take advantage of its restricted scope of application. First, it is applied only in a selected number of municipalities, thereby generating regional variation. Second, the condition of rental objects generates an additional dimension of variation since new and modernised objects are exempt from rent control. Based on data for rental offers in Germany, we apply a triple-difference framework with regionspecific time trends as well as flat type-specific ones. Despite the high political expectations, our estimates indicate that the German rent control dampens rental price by only 2.5%. This effect varies across object characteristics and seems to be larger for lower-quality dwellings and in the lower price segment. Nevertheless, the application of an event-study indicates that these effects are not persistent over time.

61 A Machine Learning-based Multivariate Non-Parametric Density Estimator

Patricia Dörr; Ralf Münnich

Universität Trier, Deutschland
doerr@uni-trier.de

SECTION: Poster

Self-organizing maps (SOMs) are successfully used in applications of machine learning such as clustering, prediction and visualization. Three of their appealing features are nearly topology preservation and dimension reduction, and they proceed without supervision. SOMs are thus closely connected to the structure of the input data and consequently a popular tool for multivariate data exploration. Furthermore, SOMs are subject of ongoing research about their theoretical properties like convergence in distribution. Nonetheless, they have rarely been applied to derive a – possibly multivariate – density estimator. In this contribution, the theoretical properties and necessary extensions of the classical SOM algorithm that are required to motivate an application in (multivariate) density estimation are discussed. Additionally, an extension to conditional probability density estimation is described. In a simulation study, the introduced estimator is compared to the traditional kernel density estimator with respect to the Kullback-Leibler divergence and the integrated mean squared error.

62 Application of location-based Data Analysis for Census Planning (Case study in Iran)

Saeed Fayyaz; Arash Fazeli

Iran's National Statistical Organization, Iran
S_fayyaz@sci.org.ir

SECTION: Poster

Despite the growing practices in location-based data and its analytics use, there is still the paucity of research on links between government location-based data analytics use and public value creation. The census agencies are at varying stages in digitally transforming their national census process, products and services through assimilating and using big data techniques. In this paper, big data techniques on location-based data of trail implementation of population and housing census on 2015 has been applied in Iran. This trial was carried out by more than 76 enumerators with Tablet devices in Iran. Regarding to the gathered data that had been registered by enumerators, a clustering analysis has been done to simulate the enumerators' tasks down by month, day, hour so as to reach a norm for progressing in different cities or sub-cities. Moreover, by this kind of analysis scrolling time and distances, mean time of completing questionnaire for eligible households, progressing trend etc. will be possible to stimulation down by any desire level including city, sub-city, rural etc. Future, planning for absent households in census as problematic issue will be easy simulated down by areas, geographical zones, weekly days, methods of enumeration etc. In order to provide these analysis, different steps including data initial recognition, cleaning and purifying, data description, times simulation, hypothesis test and conclusions have been presented. Obtained result revealed that simulation' finding are highly similar with real outcomes of implementation and this result will be used as norms and criteria for main census in Iran.

63 Dynamic Microsimulation Modelling of Health among older Migrants in Germany

Christoph Frohn; Petra Stein

Universität Duisburg-Essen, Deutschland
christoph.frohn@uni-due.de

SECTION: Poster

In the German society an increase in the quantitative relevance of the population of elderly with a migration background can be expected in the future. This raises questions about the future health situation of older migrants and the differences to the autochthonous population. In socio-political terms, knowledge about future health among migrants is important to prepare the health care system for new challenges. However, from the point of view of the social sciences the fundamental questions are how to explain health inequalities, to what extend migrants are affected and on which way migrant populations do contribute to future developments on health.

While recent research has concluded that core dimensions of social inequality are (at least indirectly) linked to the occurrence of differences in health within the society, less attention is paid to possible implications on the future development of health in Germany. Particularly regarding life in old age and migrants in the German population, it is not clear which developments are to be expected and how these can be understood more in detail.

This contribution provides a period-oriented dynamic microsimulation model on health of older migrants in Germany for a period up to the year 2050 based on data from the German microcensus. The results of the simulation show exemplarily that microsimulations are capable of forecasting health inequalities regarding migrants modelling so-called „what if“ scenarios. Thereby, it is examined on what ways the further development of health-related factors (e.g. illness or smoking) depend namely on a large number of societal mechanisms and their interplay and also what role related macrostructural changes are playing in health developments. As a result, it becomes apparent, that the macrostructural process of developments on health for elderly migrants depends on the interdependence of socio-structural and demographic events that take place on the level of individuals.

64 Sensitivity of Composite Indicators due to Non-Sampling Errors

Laura Güdemann; Ralf Münnich

Universität Trier, Deutschland
guedemann@uni-trier.de

SECTION: Poster

Composite indicators are used for performance analysis on multidimensional constructs comparing different units such as regions or countries. They are widely applied for example to support policy decision making processes or to monitor the performance development of a unit. An example of a composite indicator is the indicator on equitable and sustainable well-being by Istat which is used to monitor and compare the well-being of citizens in Italian municipalities. Its sub-indicators are estimated based on data from different surveys, such as for example EU-SILC data.

The quality of the interpretation on these comparisons highly depend on the quality of the composite indicators and its estimated sub-indicators from survey data because sub-indicators are estimated from different surveys with different quality issues. In order to be able to apply fair comparisons of the performances on the composite indicators it is necessary to include the uncertainty inherit in survey data quality issues in the analysis.

An example of a common data quality issue in survey data is the presence of missing values due to non-response. This is often handled with multiple imputation which is used in this study as an example. Using the repeated-imputation inference, a measure for the uncertainty inherit in the multiple imputation can be calculated.

The study aims to show how this measure of uncertainty can be used in the construction of the composite indicator on equitable and sustainable well-being. For the aggregation of the sub-indicators the Data Envelopment Analysis is used which makes it possible to included imprecise data, which is expressed as an interval, in the construction of the composite indicator. As a result, the composite indicator itself will be expressed as an interval, reflecting on the uncertainty which is based on the data quality issues.

65 Digitale Aktionsraumforschung. Neue Methoden zur Analyse von Bewegungsprofilen im städtischen Raum

Marc Hannappel; Philipp Jakobs

Universität Köln, Deutschland
MarcHannappel@uni-koblenz.de

SECTION: Poster

Ziel der Aktionsraumforschung ist die Erhebung und Auswertung von Aktionsräumen der städtischen Bevölkerung, d.h. die Analyse der individuellen Bewegungsmuster und der von den Personen aufgesuchten öffentlichen und privaten städtischen Räume. Seit einigen Jahren stehen neue Technologien im Bereich computergestützter Erhebungs- und Auswertungsmethoden zur Verfügung. Für die Aktionsraumforschung bieten sich in diesem Zusammenhang gerade solche Erhebungsmethoden an, die die räumliche Bewegung von Personen digital in Form von GPS-Daten erfassen und für empirische Analysen zugänglich machen können. Im Zentrum des Vortrages steht die Präsentation eines neuen Forschungskonzeptes zur Erhebung und Auswertung digitaler Aktionsräume. Zu diesem Zweck haben wir innerhalb eines Mixed-Methods-Ansatzes quantitative und qualitative Erhebungs- und Auswertungsverfahren miteinander kombiniert und das Konzept in einer explorativen Studie mit Koblenzer Studierenden getestet. Inhaltlich zielt die Forschungsarbeit auf die Klärung folgender drei Fragen: (1) Lassen sich Bewegungsprofile entlang sozio-kultureller Merkmale beschreiben? (2) Lassen sich städtische Orte identifizieren, die sich in Bezug auf sozio-kulturelle Merkmale ihrer Nutzerinnen und Nutzer durch eine besonders heterogene oder besonders homogene Nutzerstruktur auszeichnen? (3) Durch welche inhaltlichen Merkmale zeichnen sich die jeweiligen Orte als erfahrbare Qualitäten aus? Entlang dieser Forschungsfragen haben wir (1) studentische Bewegungsdaten mit Hilfe einer Smartphone-App erhoben, um studentische „Hotspots“ zu lokalisieren. Zudem konnten wir (2) diese Daten mit Angaben zu den individuellen Lebensstilen der Studierenden verknüpfen und schließlich haben wir (3) die erhobenen Daten mit qualitativen Beobachtungen der lokalisierten „Hotspots“ ergänzt, um ein näheres Verständnis für die Hintergründe der Bewegungsmuster zu entwickeln. Ohne Anspruch auf Repräsentativität erheben zu können, zeigen unsere Ergebnisse, dass sich die Bewegungsprofile bzw. die identifizierten Lokalitäten entlang der lebensstilspezifischen Differenzierungskriterien beschreiben lassen. Werden die jeweiligen Orte mittels qualitativen Beobachtungen näher betrachtet, dann lassen sich Anzeichen einer engen Passung von theoretisch implizierten Annahmen zu den Lebensführungstypen und den Raumbezügen einerseits und den aufgesuchten Lokalitäten andererseits feststellen.

66 Der Einfluss berufsstruktureller Informationen auf Stellenbesetzungsprozesse — heute und in der Zukunft?

Tobias Maier (1); Stefanie Steeg (1); Gerd Zika (2)

1: Bundesinstitut für Berufsbildung, Deutschland; 2: Institut für Arbeitsmarkt und Berufsforschung
tobias.maier@bibb.de

SECTION: Poster

In Zeiten des Fachkräftemangels ist es unerlässlich, die Ursachen von Rekrutierungsschwierigkeiten für Betriebe zu verstehen. Es gibt jedoch verschiedene Gründe, warum Unternehmen nicht ausreichend Fachkräfte für die Besetzung ihrer Stellen finden können. Einige sind unternehmensspezifisch, andere können sich auf bestimmte Merkmale des Berufs beziehen, da der deutsche Arbeitsmarkt nach Berufen strukturiert ist. So zeigt sich in den Vakanzzeiten der Bundesagentur für Arbeit, dass einige Berufe seit Jahren dauerhaft überdurchschnittlich hohe Vakanzzeiten aufweisen, während in anderen Berufen keine Rekrutierungsschwierigkeiten bestehen.

Wir untersuchen, welchen Erklärungsbeitrag berufsspezifische Charakteristiken für die durchschnittlich abgeschlossenen Vakanzzeiten der BA in den Jahren 2012 bis 2015 liefern können. Dieses Vorgehen verfolgt zwei Ziele: Zunächst möchten wir beurteilen, inwieweit die berufsspezifische Rekrutierungssituation mit berufsspezifischen Merkmalen zusammenhängt, um entsprechende politische Maßnahmen benennen zu können, welche die Rekrutierungssituation beeinflussen könnten. In einem zweiten Schritt geben wir einen Ausblick auf die erwartete Entwicklung der berufsspezifischen Rekrutierungssituationen, indem wir die berufsspezifischen Indikatoren mit einer Projektion des Arbeitsangebotes und der Arbeitsnachfrage nach Berufen verbinden www.qube-projekt.de.

In unserer Präsentation werden wir insbesondere drei berufsspezifische Indikatoren diskutieren. Der erste Indikator ist ein Volumenindikator (VI), der die Höhe der von den Erwerbspersonen angebotenen Arbeitsstunden berücksichtigt. Der Qualifikationsindikator (QI) vergleicht die Nachfrage nach Erwerbstätigen im Beruf mit dem berufsspezifisch qualifizierten Angebot aus dem Bildungssystem. Der Substitutionsindikator (SI) vergleicht die Zahl der Erwerbstätigen in einem Beruf mit einem berufsspezifischen Abschluss mit der Zahl aller Erwerbstätigen im Beruf. Darüber hinaus berücksichtigen wir die Struktur des berufsspezifischen Arbeitsangebots (z.B. Alter, Geschlecht) und den im Beruf gezahlten Lohn. Der Anteil an sozialversicherungspflichtig Beschäftigten im Beruf dient uns als Proxy für den Einschaltungsgrad der BA bei Stellenbesetzungsprozessen.

Da wir die zukünftige, potenzielle Vakanzzeit für einen Beruf prognostizieren möchten, verwenden wir ein parametrisches Überlebensmodell mit einer log-logistischen Verteilung der Überlebenszeit um den Einfluss der Kovariaten darzustellen.

67 Identifizierung von Ausbildungsenden ohne Abschluss in der IEBS

Caroline Neuber-Pohl

Bundesinstitut für Berufsbildung, Deutschland
neuber-pohl@bibb.de

SECTION: Poster

Die Forschung und Indikatorik über Ausbildungsabbrüche sind insbesondere durch die Datenlage begrenzt. Individualdaten zu Ausbildungsverläufen in Deutschland beschränken sich oft auf relativ kleine Stichproben, sind kaum als Panel verfügbar, oder der allgemeinen wissenschaftlichen Community nicht zugänglich. Die Stichprobe der Integrierten Erwerbshistorien (IEBS) des Instituts für Arbeitsmarkt- und Berufsforschung (IAB) könnte diese Lücke schließen, bringt jedoch beträchtliche Herausforderungen für die akkurate Identifikation von Ausbildungsenden mit sich. Aus diesem Grund, soll in diesem Beitrag die Nutzbarkeit der IEBS zur Analyse von Ausbildungsabbrüchen diskutiert und ein Lösungsvorschlag vorgestellt werden.

Die IEBS bietet mehrere Vorteile. Zum einen ermöglicht sie durch ihre Datenstruktur Panel- und Ereigniszeitanalysen. Des Weiteren, bietet sie Informationen zum Ausbildungsbetrieb insbesondere auch über die Verlinkungsmöglichkeiten zum IAB Betriebspanel. Außerdem bietet die große Zahl der abbildbaren Ausbildungsverläufe in der IEBS über einen großen Zeitraum insbesondere auch Repräsentativität für kleinere Berufssegmente. Gerade dadurch, dass der Dimension des Berufs im Kontext von Ausbildungsverläufen eine wichtige Rolle zukommt, besteht hierin ein wesentlicher Vorteil gegenüber den existierenden Befragungsdaten.

Im Gegensatz zur Berufsbildungsstatistik z.B. sind in der IEBS jedoch Ausbildungsenden und Ausbildungserfolg nicht direkt, sondern nur über die Veränderung im Erwerbsstatus erkennbar. Der Ausbildungserfolg erschließt sich wiederum nur über die Veränderung im durch den Betrieb freiwillig mitgeteilten Bildungsabschluss. Hierdurch sind erfolglose Ausbildungsenden nur sehr unsauber erfassbar. Im Rahmen eines Forschungsprojekts zu den Determinanten von erfolglosen Ausbildungsenden wurde eine Strategie erprobt, mit den Herausforderungen der Daten umzugehen und Ausbildungsenden, die nicht zu einem Abschluss führen, in den Daten zu identifizieren. Hierzu wird insbesondere eine auf die Identifikation von Ausbildungserfolg angepasste Bereinigungsroutine der Bildungsabschlüsse vorgeschlagen. Des Weiteren wird ein Verfahren getestet, welches nicht angezeigte erfolgreiche Übernahmen durch Lohnsprünge identifiziert, die den jeweiligen Lohnunterschieden zwischen Auszubildenden und qualifiziertem Personal entsprechen.

68 Developing a Spatial Microsimulation Model for Labor Market Integration of Migrants in Germany

Monika Obersneider; Petra Stein

University of Duisburg-Essen, Germany
monika.obersneider@uni-due.de

SECTION: Poster

Microsimulation modeling has gained growing attention in the last years as a useful statistical tool for individual based population forecasts. Yet, only a handful datasets that are used for microsimulations allow for modeling with respect to the small level spatial structure the individuals are embedded in. Hence, the interplay of regional disparities with individual and contextual outcomes has received insufficient attention in this field. The aim of this contribution is to give an insight into the usage of a nationwide spatial microsimulation infrastructure in Germany („MikroSim“). We will present its application through modeling labor market outcomes of migrant populations. The resulting model can be used for forecasting. Special features of this forecast will be the integration of regional small area estimates as well as the derivation of estimates from empirical results. This allows for the simulation of different „what if“ scenarios and future dynamics of the labor market integration of ethnic minorities on several societal levels of interest.

69 The (Non-)Significance of Reporting Errors in Economics: Evidence from Three Top Journals

Peter Pütz (1); Stefan B. Bruns (2)

1: Georg-August-Universität Göttingen, Germany; 2: University of Hasselt, Belgium
ppuetz@uni-goettingen.de

SECTION: Poster

Coefficients, standard errors, test statistics and p-values constitute the core output in empirical economics and empirical research in general. These statistical values are essential for the cumulative research process and frequently used in evidence-based decision making. Therefore, it is of eminent importance that these statistical values are reported correctly. Moreover, erroneous results would question the thoroughness of researchers and the quality of the peer-review process in academia.

We investigate the prevalence and sources of reporting errors in hypothesis tests in three top economic journals. Reporting errors are defined as inconsistencies between reported significance levels and statistical values such as coefficients and standard errors. We analyze 30,993 tests from 370 articles and find that 33% of the articles contain at least one reporting error. Survey responses from the respective authors, replications and regression analyses suggest some simple solutions to mitigate the prevalence of reporting errors in future research. Open data and software code policies in line with a vivid replication culture seem to be most important.

70 Benefit underreporting in survey data and its consequences for measuring non-take-up: new evidence from linked administrative and survey data

Kerstin Bruckmeier (1); Regina T. Riphahn (2); Jürgen Wiemers (1)

1: Institut für Arbeitsmarkt- und Berufsforschung (IAB), Germany; 2: Friedrich-Alexander Universität Erlangen-Nürnberg
Kerstin.Bruckmeier@iab.de

SECTION: Microsimulations – Applications and Methods

The international literature studies non-take-up behaviour of eligible populations to evaluate the effectiveness of government programs. A challenge in the empirical analysis of non-take-up is its measurement. A correct measurement of non-take-up requires valid information on program eligibility and take-up. Most studies must rely on survey data to measure both eligibility and program participation. A major challenge in this literature is the measurement error regarding benefit take-up if the information provided by survey systematically differs from true values. Respondents might misreport benefit receipt for several reasons, e.g., they could misremember the name or timing of benefit receipt. In this study, we link survey data to administrative data, which informs us about the true benefit take-up of survey respondents.

We consider a general income support program that is available for the working-age population in Germany (Arbeitslosengeld II, Unemployment Benefit II (UB II)). We simulate eligibility for UB II based on data from the Survey „Labour Market and Social Security“ (PASS), a panel study representative for the German population. Our simulation results in a non-take-up rate of 40 percent with substantial heterogeneity across subgroups, which is in line with findings based on other data. We use our administrative data on actual UB II receipt to correct for benefit underreporting in the survey and to reclassify non-take-up outcomes. This reduces the overall take-up rate from approximately 40 percent to approximately 35 percent. Next, we investigate how benefit misreporting biases the estimates of a standard model of the take-up decision among eligible households. Our findings suggest that correcting for underreporting affects not only the level of non-take-up but also the impact of the correlates of the take-up decision as the extent of measurement varies with the characteristics of the misreporting population and is associated with the drivers of underreporting.

71 Adjusting the welfare system to new labour market risks: Integrating an in-work benefit into the German basic income system

Kerstin Bruckmeier; Jannek Mühlhan; Jürgen Wiemers

Institut für Arbeitsmarkt- und Berufsforschung, Deutschland
jannek.muehlhan2@iab.de

SECTION: Microsimulations – Applications and Methods

Labour market risks in Germany have increasingly shifted from unemployment towards unstable low-paid jobs and in-work poverty. This paper proposes an adjustment of the German welfare system to these new labour market risks by implementing an in-work benefit (IWB) based on the rules of the existing basic

income system. The aim of the reform is to reduce the complexity of the benefit system, raise incentives to work full-time, increase benefit take-up, and reduce poverty. We employ the microsimulation model of the Institute for Employment Research (IAB-MSM) to analyse fiscal, distributional, and labour supply effects of the proposed IWB. Our results show that the reform can achieve significant effects on labour supply and poverty reduction compared to recently implemented reforms of the government

72 REGIONAL MICROSIMULATION MODEL OF THE FUTURE NEED FOR PRIMARY SCHOOL INFRASTRUCTURE IN TRIER

Sebastian Dräger

Trier University, Deutschland
draegers@uni-trier.de

SECTION: Microsimulations – Applications and Methods

There are only few practical examples in which methods of microsimulation have been used to analyze the provision of a certain public infrastructure in different sectors to anticipate long-term conditions. The usage of a spatial perspective that allows taking regional disparities and local characteristics – such as specific settlement structures – into consideration, is even more uncommon. An example for spatial microsimulation models, in which not only the possibility of a forward projection but also the spacial perspective are highly important, is the future demand for primary schools and the consequential need for trained teaching staff.

In this paper, we show how a microsimulation study in Trier was conducted to further explore if spatial microsimulations can serve as an instrument to analyze the future need of elementary school infrastructure. The basis of this study is a synthetic dataset of Trier's population that has been created based on the results of the German Census 2011. The population was located in the municipal area using Census grid cells. In order to validate this synthetic dataset and the various models of individual behavior that determine the simulation process, the results of this models have been compared with precise data about the numbers of primary-school pupil from 2011 to 2018.

Even simple models of the included simulation parameters show that the synthetic population and the forward projection of it can illustrate the current situation in Trier. In combination with the possibility to observe the distribution of the pupils in the city area and its future development, this study displays that spatial microsimulation models can pose as a tool to analyze and plan future supply situations, even though certain limitations to this procedure also become visible.

73 Building a longitudinal income module for microsimulations of spatial inequality

Jana Emmenegger (1); Ralf Münnich (2)

1: Destatis, Germany
jana.emmenegger@destatis.de; 2: Trier University, Germany

SECTION: Microsimulations – Applications and Methods

This paper examines a novel administrative data source for microsimulations: Income modules for microsimulation models are usually based on survey samples. This is associated with various methodological challenges and requires extensive adjustments. For the first time, the German taxpayer panel .- an income register recording the total population of more than 54 million taxpayers in Germany from 2001 to 2014 – is used for dynamic, spatial microsimulations. The data set reveals the full income distribution in Germany and therefore provides an ideal basis to analyse income and its development over time.

The aim is to develop an income module for a large German microsimulation model called MikroSim that is currently being built jointly by the universities of Trier and Duisburg-Essen in cooperation with the Federal Statistical Office. This will provide necessary information to enable a variety of microsimulations in general and particularly for income-related policy analysis, such as tax or family policy reforms. Moreover, income also serves as an important explanatory variable for other modules, such as fertility decisions, internal migration or education opportunities of children.

Mixed model approaches are employed to estimate individual income and the difference in incomes over time. Precisely speaking, different generalized additive models with structured and unstructured spatial effects as well as p-splines are estimated. Where necessary, income disparities over time are modeled following a Markov process. As an alternative methodological approach, in a later stage, it shall be compared whether the use of structured additive distributional regression further enhances predictive performance of income estimates for microsimulations by providing estimates of the full income distribution conditional on the included covariates. Exemplary results of spatial income inequality in Germany are presented. The research is developed within the research group MikroSim (FOR 2559), funded by the German Research Foundation.

74 30 years environmental metrics – What have we learnt? Where to go?

Walter J. Radermacher

La Sapienza University Rome, Italy
wjr@outlook.de

SECTION: Environmental Statistics

Twenty-two years after the 51st Session of the International Statistical Institute in Istanbul in 1997, many of the statistical questions and answers on the topic of Sustainable Development already presented by the author at the time are still valid: Sustainable Development can only be defined and achieved by a complicated restructuring process of the society. Consequently, environmental statistics must be interpreted as representations of possible margins for the manoeuvre to Sustainable Development, and inputs to policy debate in this sense. With the UN Sustainable Development Goals, these questions have gained enormous political significance. New techniques and data sources open up completely new possibilities for statistics. At the same time, however, environmental problems have worsened, especially at the global level, so that the development of methods and the production of solid statistics must be driven forward with great time pressure.

If it is however not possible to measure Sustainable Development (and quantify progress in this regard) independently of the political-social change process, the statistical question behind it is with which metrics the decisions on the way to Sustainable Development can be improved and of what quality and condition these metrics have to be, so that they are robust enough for the social conflicts that have to be solved. The complexity inherent in Sustainable Development must be reduced and simplified, but without going too far or risking the credibility of metrics with covert value judgments. Indicators are a very suitable tool for this. However, for indicators to actually play their intended enlightening role, more is needed than statistical methodology and reliable production of information. In addition, factors must be taken into account which, both in the design and in the communication of indicators, place their functionality and the target group of users in the political-societal context in the foreground.

75 Climate Change Detection and Attribution: An applied problem in Environmental Statistics

Andreas Hense

University Bonn, Germany
ahense@uni-bonn.de

SECTION: Environmental Statistics

Climate Change detection is defined as the demonstration that „climate has changed in some defined statistical sense without providing a reason for that change“. Attribution is defined as „the process of evaluating the relative contributions of multiple causal factors to a change or event with an assignment of statistical confidence“. (citations from the IPCC assessment report No. 5). Various sources of uncertainty

like the error structure of the observations, the error structure of the inherent climate noise of the global climate model simulations and the epistemic errors of the used complex climate models arise. An approach is offered by Bayesian decision analysis. We will describe an implementation which makes use of the simulation data sets offered by the coupled model intercomparison (CMIP) project in its phase 3 and 5 and available observational data sets for air temperatures. Assuming for likelihood and prior distribution multivariate Gaussians the posterior probability of a specific scenario (no changes of external climate forcings, natural (observed solar, volcanic) changes of external forcing, anthropogenic changes (observed greenhouse gases, sulfate aerosols) of external forcing, all changes of climate forcing within the 20th century) given the observations can be calculated by analytic integrations. Full rank precision matrices are determined with the glasso algorithm, the problem associated with the multi model „ensemble of opportunity“, which assumes that the CMIP model sample is i.i.d. , is studied by Gaussian mixture modelling for the individual models. The potential scale mismatch between observations and simulations is tackled by maximizing the posterior using regression factors applied to the simulations. Results using global mean near surface temperatures between 1900 and 2010 as well as radiosonde observations over Germany 1961-2010 will illustrate the outcome of the climate change detection and attribution study.

76 Return times of extreme weather events

Katharina Hees

TU Dortmund, Germany
hees@statistik.tu-dortmund.de

SECTION: Environmental Statistics

Not only the strength of extreme weather events but also the questions: „How often have we to fear its appearance?“ or „How likely is it that there will be an extreme weather event in a fixed period?“ are of great importance. The traditional approach is to assume that a Poisson process can model the occurrence of the extreme events, which means that the times between subsequent exceedances are exponentially distributed. The underlying time series often displays short-term dependence, which leads to a temporal clustering of extreme events. Nowadays, inference is then commonly based on the identification of independent clusters by some declustering mechanism, assuming the occurrence of clusters to follow a Poisson process with exponential recurrence times in between subsequent clusters. However, not only the assumption of serial correlation in the underlying time series but also heavy-tailed inter-arrival times or dependencies in the arrivals can lead to clustering of extremes and hence to different distributions of the return times. In this talk, I will give a short overview of different approaches, talk about how to model the inter-arrival times between extreme weather events and present ideas on how to choose an appropriate model.

77 Spatial Modeling of Heavy Precipitation in the South of France Using Max-Stable Coupling Between Weather Station Recordings and Ensemble Forecasts

Marco Oesting (1); Philippe Naveau (2)

1: University of Siegen, Germany; 2: LSCE-CNRS, France
oesting@mathematik.uni-siegen.de

SECTION: Environmental Statistics

Due to complex physical phenomena, the distribution of heavy rainfall events is difficult to model in space at fine scales. In order to capture effects such as non-stationarity and anisotropy, in this talk, we propose to couple different sources of data on extremes. More precisely, we construct data driven mathematically sound processes by coupling forecast ensemble data from the French national weather service (Météo-France) with local observations. The spatial processes we provide are parsimonious in parameters, easy to simulate and capable of incorporating nugget effects and reproducing spatial non-stationarities.

78 The Rotated Empirical Similarity Approach: Theory and Applications

Jamol Bahromov; Vasyl Golosnoy; Markus Pape

Ruhr Universität Bochum, Germany
jamol.bahromov@rub.de

SECTION: Empirical Economics

The empirical similarity (ES) approach is a comparatively new econometric technique suitable to recover the mechanism behind experience-based decisions on the variable of interest (Gilboa et al., 2006, 2011). The ES roots in the case-based decision theory (cf. Gilboa and Schmeidler 2001) which postulates that the current decision should be close to successful decisions in similar past situations.

Hence, the ES amends rule-based decision principles which rely on some formal rules by decisions related to case-based reasoning which are primarily based on experience.

The estimation of ES allows to infer from empirical data how decision makers evaluate the similarity of current and past situations based on the vector of exogenous variables characterizing the situation or case (Lieberman 2010). The ES approach is also suitable for prediction purposes (Golosnoy et al., 2014).

In this paper we extend the ES approach as in Gilboa et al. (2006, 2011) by introducing the possibility to apply a rotation to the exogenous characteristics in order to assess the similarity of situations more completely, including interaction effects between the exogenous variables. We provide the theory behind this rotated ES (RES) approach and propose a novel test in order to differentiate between an informative rotation and a mere sign permutation, which does not exploit interaction effects. Finally, we illustrate our RES concept in two empirical applications analyzing hourly day-ahead prices for electricity and the number of housing starts for private dwellings in the U.S.

79 Estimating the transition time from school to university using a stochastic mortality model

Britta Stöver

Institut für Statistik an der Leibniz Universität Hannover, Deutschland
stoever@statistik.uni-hannover.de

SECTION: Empirical Economics

The provision of university places for first year students facilitates the participation in education as well as the accumulation of human capital. The higher level of education achieved by university education supports participation opportunities in the labour market and income prospect so that a precarious living condition in old age becomes less likely. Additionally, a huge number of empirical studies give evidence that the accumulation of human capital has a positive impact on GDP.

Thus, the provision of a sufficient number of university places meeting the demand of school leavers who want to enrol in university is of high sociopolitical importance. An economic model forecasting the demand for university places in each Federal State in Germany until 2030 is supposed to provide

estimation about future demand and hence offer more planning security. One important model component is the transition from school to university. There is no time limit for the gap between the achievement of the university admittance qualification and the enrolment into university and many different aspects influence the transition process. Nevertheless, most school leavers start studying immediately or one year after graduating from school. Additionally, a tendency towards a shortening of the transition time can be observed in the past. In the current model version however it was assumed that the transition behaviour remains unchanged for the future. Based on a stochastic mortality model this assumption was relaxed and the impact on the model results assessed: The transition rates were interpreted as mortality rates that were estimated using the Lee-Carter model and projected based on an ARIMA process. The results suggest that for the majority of Federal States the assumption of constant transition rates proved to be indeed the best solution.

80 Effectiveness and Efficacy of R&D Subsidies: Estimating Treatment Effects with One-sided Noncompliance

Philipp Böing (1,2); Bettina Peters (1,3)

1: ZEW - Leibniz-Zentrum für Europäische Wirtschaftsforschung, Germany; 2: Peking University; 3: University of Luxembourg
b.peters@zew.de

SECTION: Empirical Economics

In evaluating the effectiveness of R&D subsidies, the literature so far has completely neglected the possibility of misappropriation of public funds. This paper contributes to the literature by evaluating the causal effect of R&D subsidies on R&D expenditures when monitoring is weak and misappropriation takes place due to moral hazard behavior. Our analysis is based on Chinese firm-level data for the period 2001-2011. Misappropriation is a major concern as we calculate that 42% of grantees misused R&D subsidies, corresponding to 53% of the total amount of R&D subsidies. In a setting with one-sided noncompliance to funding contract rules, we differentiate between the intention-to-treat (ITT) effect and the complier average causal effect (CACE). The ITT shows how effective the R&D policy was in practice when misappropriation exists. The CACE, in contrast, depicts how effective the policy could have been without misappropriation and thus is a measure for the efficacy of the R&D subsidy policy. Combining entropy balancing and IV methods to estimate both ITT and CACE, the ITT results show mild partial crowding out of R&D expenditures. Most strikingly, however, the CACE turns out to be more than twice as large as the ITT and confirms additionality of R&D subsidies. Thus, misappropriation of R&D subsidies considerably undermines the efficacy of Chinese R&D programs.

81 Random Forest Estimation of the Ordered Choice Model

Michael Lechner; **Gabriel Okasa**

University of St.Gallen, Switzerland
gabriel.okasa@unisg.ch

SECTION: Empirical Economics

In econometrics so-called ordered choice models are popular when interest is in the estimation of the probabilities of particular values of categorical outcome variables with an inherent ordering, conditional on covariates. In this paper we develop a new machine learning estimator based on the random forest algorithm for such models without imposing any distributional assumptions. The proposed Ordered Forest estimator provides a flexible estimation method of the conditional choice probabilities that can naturally deal with nonlinearities in the data, while taking the ordering information explicitly into account. In addition to common machine learning estimators, it enables the estimation of marginal effects as well as conducting inference thereof and thus providing the same output as classical econometric estimators based on ordered logit or probit models. An extensive simulation study examines the finite sample properties of the Ordered Forest and reveals its good predictive performance, particularly in settings with multicollinearity among the predictors and nonlinear functional forms. An empirical application further illustrates the estimation of the marginal effects and their standard errors and demonstrates the advantages of the flexible estimation compared to a parametric benchmark model.

82 Einkommensungleichheit zwischen den bayerischen Kreisen 1991 bis 2016

Miriam Orlowski

Bayerisches Landesamt für Statistik, Deutschland
miriam.orlowski@statistik.bayern.de

SECTION: Regional Statistics

Zwischen den bayerischen Landkreisen und kreisfreien Städten bestehen beträchtliche Differenzen in den verfügbaren Einkommen pro Kopf. In folgendem Artikel untersuchen wir die Einkommensungleichheit in Bayern deskriptiv anhand gängiger Ungleichheitsmaße. Zielgrößen sind dabei die verfügbaren und primären Einkommen der privaten Haushalte im Zeitraum 1991 bis 2016. Insgesamt nahm die Ungleichheit der Einkommen in den letzten 25 Jahren ab, was maßgeblich auf eine große Zahl an Kreisen zurückzuführen ist, die an Einkommen aufgeholt haben, darunter vor allem Kreise in Niederbayern und der Oberpfalz.

83 Eine Faktoren- und Clusteranalyse der Raumordnungsregionen Deutschlands auf Basis von Projektionsdaten des Modells QMORE für den Zeitraum 2015-2035

Frederik Parton

Gesellschaft für wirtschaftliche Strukturforchung mbH, Deutschland
parton@gws-os.com

SECTION: Regional Statistics

Ziel der Untersuchung ist es, die Unterschiede in der Entwicklung der ökonomischen Wettbewerbsfähigkeit der unterschiedlichen Regionen Deutschlands für den Zeitraum 2015 bis 2035 darzustellen. Kern der Untersuchung stellen eine Faktoren- sowie eine Clusteranalyse dar. Diese multivariaten statistischen Methoden werden dazu genutzt, die 96 Raumordnungsregionen Deutschlands unter Berücksichtigung von unterschiedlichen sozioökonomischen Indikatoren zu möglichst homogenen Gruppen (Cluster) wirtschaftlicher Leistungsfähigkeit zusammenzufassen. Die Cluster werden für unterschiedliche Zeitpunkte des Projektionszeitraums 2015 bis 2035 berechnet und kartografisch dargestellt. Als Ergebnis wird die Entwicklung der Cluster im Zeitverlauf und somit die Entwicklung der ökonomischen Wettbewerbsfähigkeit der Regionen Deutschlands präsentiert.

Als Datenbasis dienen Projektionen des Modells QMORE (Qualifikationen Monitoring Regional) für den Zeitraum von 2015 bis 2035. Das Modell QMORE wurde von der Gesellschaft für wirtschaftliche Strukturforchung (GWS) mbH, in Zusammenarbeit mit dem Institut für Arbeitsmarkt- und Berufsforschung (IAB) sowie dem Bundesinstitut für Berufliche Bildung (BIBB) im Rahmen des QuBe-Projektes (www.qube-projekt.de) entwickelt.

84 Regionale Disparitäten in Niedersachsen – dargestellt anhand einer Clusteranalyse auf Gemeindeebene

Arne Lehmann

Landesamt für Statistik Niedersachsen, Deutschland
arne.lehmann@statistik.niedersachsen.de

SECTION: Regional Statistics

Zur Identifizierung regionaler Disparitäten hilft in der Regionalstatistik das Instrument der Clusteranalyse. Im vorliegenden Fall wurden die Einheits- und Samtgemeinden des Landes Niedersachsen ausgehend von einer strukturbestimmenden Datenauswahl mittels mathematischen Verfahrens (Ward-Verfahren) in verschiedene Gruppen (Cluster) aufgeteilt. Die Indikatoren repräsentieren sieben Themengebiete: Demografie, Wirtschaftskraft, Arbeit, Bildung, Soziales und Wohlstand sowie Öffentliche Finanzen. Die Cluster sollten in sich homogen sein und untereinander möglichst heterogen sein.

Im Ergebnis zeigt sich, dass die Bevölkerungsdichte zwar wichtige Erklärungszusammenhänge zu den Strukturen liefert. Sie ist aber nicht alleinig maßgeblich. Genauso haben größere Städte eine bedeutende „Strahlkraft“ für das Umland. Abseits der größeren Städte gibt es jedoch auch ländliche strukturstarke Regionen. Auffällig ist in Niedersachsen eine räumliche Konzentration dieser ländlichen Gebiete im Westen des Landes. Klein- und Mittelstädte weisen überwiegend durchschnittliche Werte auf.

Die Analyse hat ausschließlich Ist-Zustände betrachtet. In der anschließenden Betrachtung der Entwicklung zeigte sich bereits bei drei Indikatoren eine Verfestigung der strukturbestimmenden Merkmale. Erstens: Dort, wo die sozialen Problemlagen vergleichsweise groß waren, haben diese auch zugenommen. Abgenommen haben sie dort, wo sie zuvor niedrig waren. Zweitens: Die Alterung der Bevölkerung verstärkte sich in den Clustern, die schon in der Ausgangslage als „überaltert“ gelten können. Drittens: Der Arbeitsplatzbesatz stieg überall an. In den bereits vergleichsweise besser aufgestellten Clustern entwickelte er sich jedoch besser als in den weniger gut aufgestellten Clustern.

Der Vortrag basiert auf der im Statistischen Landesamt Niedersachsen erstellten Clusteranalyse aus dem Jahre 2018 und zeigt darüber hinaus die Entwicklung anhand aktueller Daten auf: Wie haben sich die Werte in den Clustern seither verändert? Entfernen sich die ohnehin strukturell schwächer aufgestellten Cluster noch weiter vom Durchschnitt und driften die Cluster weiter auseinander oder gibt es eine Angleichung?

85 Ausprägung und Veränderung regionaler Disparitäten — eine Analyse mittels Markov-Ketten

Antonia Milbert

Bundesinstitut für Bau-, Stadt- und Raumforschung, Germany
antonia.milbert@bbr.bund.de

SECTION: Regional Statistics

Markov-Ketten sind ein nützliches Tool, um regionale Disparitäten zu untersuchen. Ihr Vorteil liegt darin, dass kaum Annahmen zu berücksichtigen sind. Vor allem können über sie auch die Veränderungen von Disparitäten mittels Statistiken untersucht werden, die in ihrem Zeitverlauf gewisse Datenbrüche aufweisen. Des Weiteren sind umfassende Betrachtungen der Übergangswahrscheinlichkeiten zwischen Perzentilen möglich, die differenziertere Auskünfte über die Entwicklung von Disparitäten geben als reine Konvergenz- oder Divergenzmaße.

In dem Beitrag sollen anhand verschiedener Standardindikatoren die regionalen Disparitäten in Deutschland vorgestellt und in ihrer Entwicklung seit Mitte der 1990er Jahre dargestellt werden.

86 Quantitative Robustheit von Prädiktoren auf Basis lokaler SVMs

Florian Dumpert (1,2)

1: Statistisches Bundesamt, Germany; 2: Universität Bayreuth, Germany
florian.dumpert@destatis.de

SECTION: Nonparametric and Robust Statistics

Support Vector Machines (SVMs) – als solche eine Methode des statistischen, maschinellen Lernens – werden in der Regel zu den nicht-parametrischen Verfahren gezählt, da sie im Idealfall keine Annahmen an die zugrundeliegende Verteilung der erklärenden und der zu erklärenden Variablen treffen müssen, um wünschenswerte (asymptotische) statistische Eigenschaften zu garantieren. Man kann zeigen, dass SVMs in der Lage sind, gute Vorhersagen für die Ausprägung der zu erklärenden Variable für neue Datenpunkte zu liefern. Darüber hinaus kann man beweisen, dass SVMs robust gegenüber leichten Veränderungen der zugrunde liegenden Verteilung sind. Im Rahmen der Statistischen Woche 2017 wurden bereits erste Robustheitseigenschaften von Prädiktoren auf Basis lokaler Support Vector Machines vorgestellt. Die Nutzung mehrerer lokaler SVMs anstelle einer globalen stellt einen Ausweg für das Problem dar, dass SVMs in der Regel sehr rechenintensive Verfahren sind und somit Schwierigkeiten bei größeren Datensätzen zeigen. Der aktuelle Vortrag geht genauer auf technische Details der quantitativen Robustheit dieser Prädiktoren ein und veranschaulicht die Resultate graphisch.

87 K-sign depth: Asymptotic distribution and efficient computation

Dennis Malcherzyk; Kevin Leckey; Christine H. Müller

TU Dortmund, Germany
dennis.malcherzyk@tu-dortmund.de

SECTION: Nonparametric and Robust Statistics

K-sign depths describe the fit of a given parameter in a regression model by counting all ordered K-tupels of the residuals with K-1 sign changes. We discuss an asymptotic equivalent representation of the K-sign depth based on a functional of the random walk associated to the signs of the residuals. This representation will be used to derive the asymptotic distribution under the null hypothesis by applying a functional central limit theorem (Donsker's Invariance Principle) to construct an asymptotic statistic test. Furthermore, the high computational effort of the K-sign depth can be reduced by this representation from the asymptotic derivation. We are able to compute K-sign depth in linear time for $K < 6$ exactly and for higher K, we have the option to calculate an asymptotic version of K-sign depth in linear time.

88 Semi-parametric and robust estimation in intensity-based models

Mirko Alexander Jakubzik

Technische Universität Dortmund, Germany
jakubzik@statistik.tu-dortmund.de

SECTION: Nonparametric and Robust Statistics

The theory of counting processes provides a general framework suitable for numerous applications in engineering, epidemiology or econometrics. The cumulative intensity process corresponding to a counting process serves as a predictor and represents the process's qualitative behaviour. We introduce semi-parametric load sharing models based on intensity processes that also allow for damage accumulation. Two competing approaches to estimate the parametric part of these models are discussed:

The minimum distance estimation method of Kopperschmidt and Stute is based on a paper published in 2013 and yields consistent and asymptotically gaussian distributed estimates by minimizing the distance between the counting process and the estimated cumulative intensity. The novel maximum depth estimation method rests upon the conditional cumulative hazard transform of a point process. The proposed estimator maximizes the transform's 3-sign depth in order to obtain a more robust parameter estimation.

In a case study, the herein discussed methods are applied to fatigue test data obtained through a civil engineering experiment that took place at TU Dortmund University. Furthermore, a comparison of these methods is performed by virtue of a simulation study.

89 Public Procurement of Innovation: Evidence from a German Legislative Reform

Dirk Czarnitzki (1); Paul Hünermund (2); Nima Moshgbar (3)

1: KU Leuven, ZEW; 2: Maastricht University; 3: DICE Consult
dirk.czarnitzki@kuleuven.be

SECTION: Business and Market Statistics

The use of public procurement to promote private innovation activities has attracted increasing attention recently. Germany implemented a legal change in its procurement framework in 2009, which allowed government agencies to specify innovative aspects of procured products as selection criteria in tender calls. We analyze a representative sample of German firms to investigate whether this reform stimulated innovation in the business sector. Across a wide set of specifications – OLS, nearest-neighbor matching, IV regressions and difference-in-differences – we find a robust and significant effect of innovation-directed public procurement on turnover from new products and services. However, our results show that the effect is largely attributable to innovations of more incremental nature rather than market novelties.

90 Auftragsforschung und Innovationsfähigkeit der Unternehmen in Deutschland

Verena Eckl

SV Wissenschaftsstatistik GmbH, Germany
verena.eckl@stifterverband.de

SECTION: Business and Market Statistics

Die Generierung, Diffusion und Absorption von Wissen sind entscheidende Faktoren der Wissensgesellschaft. Unternehmen über alle Branchen- und Industriesektoren hinweg beziehen einen beachtlichen Anteil neuer Technologien aus externen, nicht unternehmenseigenen Quellen. Diese Kombination zwischen dem, aus Perspektive des Unternehmens, „neuen“ Wissen und dem im Unternehmen bereits vorhandenen „alten“ Wissen ist von zentraler Bedeutung für den Aufbau von Wissen. In Abgrenzung zu den traditionellen internen bzw. geschlossenen Innovationsstrategien wurde hierfür von Henry Chesbrough der Begriff der „offenen Innovation“ („open innovation“) geprägt. In der Statistik zu Forschung und Entwicklung im Wirtschaftssektor in Deutschland kann die Öffnung des Innovationsprozesses über die Vergabe von Forschungsaufträgen an externe Unternehmen, private und öffentliche Forschungseinrichtungen im In- und Ausland seit 20 Jahren beobachtet werden. Der Beitrag beschäftigt sich mit der Frage, inwieweit sich die forschenden Unternehmen in ihrem Innovationsprozess über die Zeit geöffnet haben und welche strukturelle Veränderungen in der Vergabe von Forschungsaufträgen in den unterschiedlichen Branchen und Größenklassen der Unternehmen in den letzten Jahren stattgefunden haben. Darüber hinaus wird untersucht, inwieweit die Vergabe von Forschungsaufträgen mit der eigenen Forschungstätigkeit und der Innovationsfähigkeit von Unternehmen korreliert ist.

91 Patentstatistik – Möglichkeiten, Verlockungen und Fehlinterpretationen

Rainer Frietsch

Fraunhofer ISI, Germany
rainer.frietsch@isi.fraunhofer.de

SECTION: Business and Market Statistics

Ein Patent ist ein temporäres, exklusives Nutzungsrecht an einer technologischen Lösung. Patente zählen seit Jahren zu den wichtigsten Innovationsindikatoren. Sie dienen als Basis zur Beurteilung der technologischen Leistungsfähigkeit oder der Technologieprofile von Organisationseinheiten (Unternehmen und Forschungsorganisationen), von Regionen oder ganzen Volkswirtschaften. Als Innovationsindikator gehören sie zu den sogenannten Throughput-Indikatoren, da sie einen Output von FuE-Prozessen und -Investitionen darstellen, gleichzeitig aber auch einen Input in weitere wirtschaftliche Umsetzungs- und Verwertungsprozesse bilden.

Patente sind sogenannte prozessgenerierte Daten. Die verfügbaren Daten umfassen im Allgemeinen die Grundgesamtheit, sie sind meist von hoher Qualität und die Analysepotenziale sind auf Grund der verfügbaren Informationen enorm. Die Daten werden in Datenbanken für einfache Analysen oder zum Rohdatenabruf angeboten – meist kostenlos oder gegen geringes Nutzungsentgelt. Dies alles macht Patente für viele (Wirtschafts-)Statistiker und Ökonomen so verlockend. Die Patentsysteme sind jedoch sehr komplex, unterscheiden sich von Regime zu Regime und auch die verfügbaren Daten sind daher nicht immer direkt vergleichbar. Sinnvolle und aussagekräftige Analysen und Interpretationen sind daher häufig komplexer als es auf den ersten Eindruck erscheinen mag.

Im Vortrag werden Patente zunächst als Innovationsindikator beschrieben und gegenüber anderen Indikatoren sowie gegenüber anderen geistigen Eigentumsrechten bzw. Schutzstrategien verortet. Relevante Aspekte von Patentanmelde-, Erteilungs- und Nutzungsprozessen werden umrissen, um dem statistischen Nutzer die Einordnung und Interpretation zu erleichtern. Es werden wesentliche Charakteristiken von Patentanmeldungen – bibliographische, prozedurale, technologie-spezifische Informationen – erläutert und sowohl Analysepotenziale wie auch Herausforderungen benannt. Darüber hinaus werden Hinweise auf Hilfsmittel (Klassifikationen, Standardisierungen, Erweiterungen) aus der quantitativen Innovationsforschung bzw. der Technometrie gegeben. Der Vortrag schließt mit einem kursorischen Überblick über Möglichkeiten der Mikrodatenanalyse, der Technologieanalyse, volkswirtschaftlicher Fragestellungen sowie Anwendungen und Erweiterungen im Kontext von Big Data.

92 Neues vom Mietspiegel – Ein Überblick zur aktuellen Diskussion aus Sicht der Statistik

Göran Kauermann (1); Michael Windmann (1); Ralf Münnich (2)

1: Ludwig-Maximilians-Universität München, Deutschland; 2: Universität Trier, Trier
michael.windmann@stat.uni-muenchen.de

SECTION: Housing

Die aufkommende Diskussion um qualifizierte Mietspiegel in Deutschland ist in den letzten Monaten deutlich wahrnehmbarer geworden. Dazu tragen das Urteil des Verwaltungsgerichts München, welches die Stadt zur Herausgabe von anonymisierten Daten verpflichtet genauso bei wie die im Koalitionsvertrag stehende Ankündigung eines neuen Mietspiegelgesetzes. Viele dieser Diskussionspunkte betreffen auch statistische Aspekte. Der Vortrag soll daher sowohl einen aktuellen Stand der Mietspiegel in Deutschland wiedergeben als auch eine Reflektion der aktuellen politischen Diskussionen.

Dabei steht zunächst die statistische Datenanalyse im Vordergrund. In der Praxis wird jedoch in zunehmendem Maße die Methode der Datenerhebung eines Mietspiegels kritisiert und beanstandet. Grundproblem bei Mietspiegeln ist, dass eine Urliste mietspiegelrelevanter Wohnungen, aus der eine Stichprobe gezogen oder die Repräsentativität der Datengrundlage überprüft werden könnte, in keiner Stadt vorhanden ist. Mit unseren Ausführungen wollen wir versuchen die Diskussion um die Datenerhebung bei Mietspiegeln voranzutreiben. Insbesondere untersuchen wir, welche unterschiedlichen Varianten der Datenerhebung und Stichprobenziehung für qualifizierte Mietspiegel in Deutschland heute existieren und gehen dabei auf die Vor- und Nachteile ein. Ebenso versuchen wir die Kosten der Methoden zu beleuchten, soweit dies aus Wissenschaftlersicht möglich ist. Die Praktikabilität der Verfahren wird ebenfalls berücksichtigt.

Weiter werden wir die unterschiedlichen Auswertungsmethoden vorstellen und darlegen, warum ein Tabellenmietspiegel (zumindest in hinreichend großen Kommunen) nicht als Standard gelten sollte. Wir werden ein klares Plädoyer für Regressionsmietspiegel geben.

Der Vortrag schließt ab mit einer Reflektion der aktuellen politischen Pläne zur Mietspiegelreform.

93 Ein Bild sagt mehr als 1000 Worte – interaktive Gemeindekarte(n) zur Bautätigkeit in Deutschland

Ingo Wagner

Statistisches Bundesamt, Deutschland
ingo.wagner@destatis.de

SECTION: Housing

Wie viele Wohnungen wurden in den Gemeinden Deutschlands im Jahr 2017 fertiggestellt? Gibt es hierbei regionale Unterschiede? Werden Wohnungen in den Ballungszentren geschaffen, um den angespannten Wohnungsmarkt zu entlasten? Wird in den Kommunen kontinuierlich Wohnraum geschaffen?

Um diese Fragen schnell und einfach beantworten zu können, wurde von den Statistischen Ämtern des Bundes und der Länder eine interaktive Gemeindekarte der Baufertigstellungen entwickelt, die jährlich aktualisiert werden soll. Der Vortrag stellt die in der Karte miteinander verknüpften Statistiken, insbesondere die Bautätigkeitsstatistiken, vor und versucht Antworten auf die oben gestellten Fragen zu geben.

Zu den im Hochbaustatistikgesetz geregelten Bautätigkeitsstatistiken gehören die Statistik der Baugenehmigungen, die Statistik der Baufertigstellungen, die Bauüberhangs- und Bauabgangsstatistik sowie die Fortschreibung des Wohngebäude- und Wohnungsbestandes. Aufgrund der Vielzahl der erfassten Merkmale lassen sich gegebenenfalls weitere Fragestellungen in weiteren Darstellungen anschaulich abbilden.

Mit der bereits veröffentlichten Gemeindekarte werden jedoch nur die aktuellen Entwicklungen dargestellt. Ebenso interessant sind aber auch Fragen die den Wohnungsbestand in den Gemeinden betreffen, wie beispielsweise: Wie viele Wohnungen je 1000 Einwohner gibt es in den Gemeinden? Was ist die durchschnittliche Wohnfläche einer Wohnung? Lebt man auf dem Land großzügiger als in der Stadt? Vor diesem Hintergrund werden im Vortrag die Überlegungen für eine weitere interaktive Gemeindekarte vorgestellt.

94 Entwicklung von regionalisierten Häuserpreisindizes in der amtlichen Statistik

Cordula Schöneich; Markus Teske

Statistisches Bundesamt, Deutschland
cordula.schoeneich@destatis.de

SECTION: Housing

Das Statistische Bundesamt veröffentlicht vierteljährlich den Häuserpreisindex, der die Preisentwicklung von Ein- und Zweifamilienhäusern sowie Eigentumswohnungen abbildet. Bislang werden neben dem Häuserpreisindex insgesamt auch Indizes getrennt nach Neubau und Bestand auf Bundesebene publiziert. Geplant ist es ab Ende Juni zudem Preisindizes für die siedlungsstrukturellen Kreistypen gemäß BBSR sowie zusätzlich für Metropolen (Top 7 Städte) zu veröffentlichen.

Datengrundlage für den Häuserpreisindex sind Transaktionsdaten aus den Kaufpreissammlungen der Gutachterausschüsse für Grundstückswerte, die quartalsweise an das Statistische Bundesamt übermittelt werden. Um Unterschieden zwischen den gehandelten Immobilien aus verschiedenen Perioden Rechnung zu tragen, wird die Methode der hedonischen doppelten Imputation eingesetzt. Dabei werden mit Hilfe von Regressionen Qualitätsunterschiede berücksichtigt.

Um eine Veröffentlichung von Indizes für Häuser und Wohnungen auf Ebene der Kreistypen zu ermöglichen, waren verschiedene Analysen und Anpassungen der bestehenden Berechnung nötig. Dies betrifft z.B. das Merkmal „Lage“ einer Immobilie oder den Umgang mit Ausreißern. Hinzu kommt, dass sich die auf weniger Datensätzen basierenden regionalen Teilindizes zum Teil sehr volatil zeigten. Für diese Fälle wurden alternative Berechnungsvarianten getestet, wie z.B. die Time-Dummy-Methode.

Im Vortrag werden einerseits die methodischen Arbeiten zur Veröffentlichung von Kreistypenergebnissen präsentiert, aber auch Hindernisse und Entwicklungspotentiale im Bereich der Immobilienpreisstatistik vorgestellt.

95 Energieverwendung in Neubau und Bestand – Gibt es Unterschiede bei den eingesetzten Energien in neuen Wohnbauten im Vergleich zum Wohnungsbestand?

Jens Dechent

Statistisches Bundesamt, Deutschland
jens.dechent@destatis.de

SECTION: Housing

Welche Beheizungsarten (Fernheizung, Zentralheizung, etc.) kommen vorwiegend bei Neubauten zum Einsatz? Wie stellt sich diese Situation im Wohngebäudebestand dar?

Welche Energiequellen dominieren bei der Beheizung und bei der Warmwasseraufbereitung in neu erstellten Wohnungen? Gibt es Unterschiede im Vergleich zum Wohnungsbestand?

Welche Bedeutung haben erneuerbare Energiequellen? Ist diese in den letzten Jahren gestiegen?

Gibt es Unterschiede bei den genutzten Energiequellen im Hinblick auf verschiedene Gebäudearten (Einfamilien- versus Mehrfamilienhäuser)?

Der Vortrag verfolgt das Ziel, auf diese Fragen Antworten zu geben. Hierzu werden die Beheizungsarten und die Energieverwendung für Heizzwecke und für die Warmwasseraufbereitung in neu erstellten Wohnungen denen im Wohnungsbestand gegenübergestellt. Als Datengrundlage werden amtliche Ergebnisse der Bautätigkeitsstatistik und der Mikrozensus-Zusatzerhebungen genutzt und erstmals miteinander verglichen.

Kernstück der amtlichen Bautätigkeitsstatistik stellen die Statistiken der Baugenehmigungen und der Baufertigstellungen dar. In der Bautätigkeitsstatistik werden genehmigungspflichtige oder ihnen gleichgestellte Baumaßnahmen erfasst, bei denen Wohn- oder Nutzraum geschaffen bzw. verändert wird. Mit der Statistik der Baufertigstellungen wird die vollzogene Hochbautätigkeit dokumentiert. Zum Erhebungsprogramm der Bautätigkeitsstatistik zählen unter anderem auch Abfragen zur Beheizungsart und zur verwendeten Energie für Heizzwecke und für die Warmwasseraufbereitung. Diese Angaben dienen als Grundlage für die Analyse der Energieverwendung bei Neubauten.

Die Mikrozensus-Zusatzerhebungen ergänzen und aktualisieren die als Totalzählungen durchgeführten Gebäude- und Wohnungszählungen. Befragt werden die Haushalte zu ihrer Wohnsituation. Die Mikrozensus-Zusatzerhebungen liefern wichtige Informationen zum Bestand und zur Struktur der bestehenden Wohneinheiten. Ähnlich wie in den Bautätigkeitsstatistiken gehören zum Erhebungsprogramm der Mikrozensus-Zusatzerhebungen auch Fragen zur Beheizungsart sowie zu den verwendeten Energiearten der Beheizung und der Warmwasseraufbereitung. Daher dient die Mikrozensus-Zusatzerhebung als Grundlage für die Analyse der Energieverwendung beim Wohnungsbestand.

96 Die regionale Pflegebedürftigkeit in Deutschland: Status Quo und künftige Disparitäten

Anja Sonnenburg

Gesellschaft für Wirtschaftliche Strukturforschung, Deutschland
sonnenburg@gws-os.com

SECTION: Regional Statistics

Die Pflegebedürftigkeit in Deutschland nimmt seit einigen Jahren rasant zu. Die aktuellen Zahlen der Pflegestatistik des Statistischen Bundesamts weisen für das Jahr 2017 mehr als 3,4 Millionen Pflegebedürftige aus und damit einen deutlichen Anstieg zu den Vorjahren. Im Jahr 2008 lag die Zahl der Pflegebedürftigen bei 2,3 Millionen und auch im Jahr 2015 lag sie noch unter der 3-Millionen-Grenze. Der Bedarf an gut ausgebildeten Pflegekräften ist damit in Deutschland weiterhin sehr hoch. Die Entwicklung der Pflegebedürftigkeit ist zudem eng an die Demografie gebunden, womit sich die Höhe des Pflegebedarfs in den einzelnen Regionen Deutschland unterschiedlich darstellt.

Die demografisch bedingte Alterung der Gesellschaft wird sich künftig fortsetzen. Um die Bedeutung der Pflegebedürftigkeit in der Zukunft abschätzen zu können, sind die Entwicklung der Pflegequoten und die Zahl der Pflegebedürftigen, ihre Verteilung auf die Regionen Deutschlands sowie der Personalbedarf von großer Bedeutung. Neben der Zeichnung eines aktuellen Bildes der Pflegebedürftigkeit in Deutschlands wird auf Basis der Modellierung mit dem gesamtwirtschaftlichen INFORGE-Modell und den untergeordneten Regionalmodellen eine Abschätzung zur zukünftigen Entwicklung gegeben. Das INFORGE-Modell verfolgt dabei einen Input-Output-Ansatz und die regionalen Projektionen unterliegen der Shift-Share-Methode. Der Fokus liegt auf den entstehenden Herausforderungen: Lassen sich Orte mit hohem Bedarf identifizieren und in welchem Ausmaß ist ein weiterer Personalaufbau in der Pflegewirtschaft nötig? Der steigende Bedarf an Betreuung ist bereits heute vielerorts eine Herausforderung für die Gesellschaft und auch zukünftig müssen weitere Lösungen gefunden werden.

97 Räumliche Verteilung der Arztpraxen in Nordrhein-Westfalen – Analyse auf der Basis des georeferenzierten statistischen Unternehmensregisters

Simon Buhl; Ann-Kathrin Richter

IT.NRW Statistisches Landesamt, Deutschland
ann-kathrin.richter@it.nrw.de

SECTION: Regional Statistics

Einleitung

Die Bedarfsplanungsrichtlinie steuert die räumliche Verteilung der Arztpraxen über die Anzahl der Kassenarztsitze in den Planungsbereichen. Innerhalb eines Planungsbereichs können die Ärzte den Praxisstandort jedoch frei wählen. Dieser Beitrag untersucht die räumliche Verteilung der Arztpraxen in Nordrhein-Westfalen in verschiedenen Regionstypen und im Verhältnis zur Bevölkerung. Am Beispiel der

Stadt Essen wird die Verteilung der Praxen innerhalb eines Planungsbereichs analysiert.

Methoden

Datengrundlage ist das statistische Unternehmensregister Nordrhein-Westfalens aus dem Berichtsjahr 2016. Dies enthält in den Wirtschaftszweigen 86.21.0 und 86.22.0 Arztpraxen, die einen steuerbaren Umsatz von über 17.500€ und/oder sozialversicherungspflichtig Beschäftigte haben. Die Analysen basieren auf INSPIRE-konformen geografischen Rastern. Über die Rasterzellen können die Bevölkerungszahlen des Zensus 2011 und die Raumtypen nach Eurostat Verstädterungsgrad zugespielt werden. Die Verteilung der Arztpraxen in Essen wird auf PLZ-Ebene untersucht.

Ergebnisse

Die 16.865 im Unternehmensregister ausgewiesenen Arztpraxen konzentrieren sich analog zur Bevölkerung in den Ballungsräumen Nordrhein-Westfalens. Insgesamt ist eine Arztpraxis rein rechnerisch für durchschnittlich 939 Menschen zuständig. Eine Differenzierung nach Raumtyp zeigt, dass Arztpraxen im ländlichen Raum durchschnittlich fast viermal so viele Einwohner betreuen wie in städtischen Gebieten (2.945 vs. 753). Während allgemeinmedizinische Arztpraxen im ländlichen Raum rein rechnerisch für fast doppelt so viele Einwohner zuständig sind wie in den dicht besiedelten Regionen (4.582 vs. 2.600), sind es bei den Facharztpraxen ca. achtmal so viele (8.241 vs. 1.060). Abweichend von der Bevölkerungsverteilung konzentrieren sich die Arztpraxen in Essen in den südlichen und mittleren Stadtteilen. Diese Konzentration spiegelt das Gefälle abnehmender sozialer Benachteiligung von Nord nach Süd wieder, das im ganzen Ruhrgebiet entlang der als Sozialäquator bezeichneten Autobahn A40 zu sehen ist.

Fazit

Arztpraxen konzentrieren sich überproportional zur Bevölkerung in städtischen Regionen und in Essen auch in Gebieten mit geringer sozialer Belastung. Für eine Bewertung der Versorgungsinfrastruktur müssen die Anzahl der Ärzte sowie Erreichbarkeitsfaktoren miteinbezogen werden. Diese könnten mittels Entfernungsanalysen bei zukünftigen Auswertungen berücksichtigt werden.

98 Not coming in today – Regional firm productivity differentials and the epidemiology of influenza

Matthias Dorner; Peter Haller

Institut für Arbeitsmarkt- und Berufsforschung (IAB), Germany
matthias.dorner@iab.de

SECTION: Regional Statistics

With over four million cases in Germany every year, influenza and acute infectious diseases of the upper respiratory tract (henceforth ARD) have the highest number of reported doctor consultations. Although the direct treatment costs are comparably low, the indirect effects, due to work absence, are far more compelling. In this paper, we estimate the effect of local ARD deceases as an exogenous shock to the production factor labor and thus on levels of firm productivity. To quantify the ARD related shock on the production factor labor, we geocode maps of weekly reported flu emergence in Germany from official influenza surveillance data. Measured by the length of the influenza season in German municipalities, these

data exhibit substantial seasonal variation as well as high variance across regions. In our main analysis, we estimate firm level production functions using data from a comprehensive German firm survey. In our main regression, we analyze total factor productivity differentials and their relationship with the local influenza intensity. First results show sizeable negative effects of the ARD diseases on firm productivity.

99 Die regionalisierte QuBe-Bevölkerungsprojektion

Michael Kalinowski (1); Gerd Zika (2); Frederik Parton (3); Doris Söhnlein (2)

1: Bundesinstitut für Berufsbildung, Deutschland; 2: Institut für Arbeitsmarkt- und Berufsforschung der Bundesagentur für Arbeit; 3: Gesellschaft für Wirtschaftliche Strukturforschung mbH
kalinowski@bibb.de

SECTION: Regional Statistics

Um die langfristigen Entwicklungen sowohl auf der Nachfrage- als auch der Angebotsseite des Arbeitsmarkts abschätzen zu können, erstellen das Bundesinstitut für Berufsbildung (BIBB) und das Institut für Arbeitsmarkt- und Berufsforschung (IAB) unter Mitwirkung der Gesellschaft für Wirtschaftliche Strukturforschung mbH (GWS) regelmäßig Qualifikations- und Berufsprojektionen (www.QuBe-Projekt.de). Die QuBe-Bevölkerungsprojektion basiert auf der Bevölkerungsprojektion des vom IAB entwickelten integrierten Arbeitskräfteangebots- und Bevölkerungsmodells. Spezifika des Modells sind, dass es zum einen zwischen Deutschen und Nichtdeutschen unterscheidet und zum anderen die einzelnen Komponenten (Geburtenziffern, Überlebenswahrscheinlichkeiten, Abwanderungen sowie die Einbürgerungen) mit zeitreihenanalytischen Methoden schätzt und projiziert.

Die altersspezifischen Geburtenziffern (getrennt nach deutschen und ausländischen Frauen), die Überlebenswahrscheinlichkeiten (getrennt für Männer und Frauen), die Einbürgerungen (ebenfalls getrennt für Männer und Frauen) sowie die Abwanderungen (nach Geschlecht und Staatsangehörigkeit) werden mit Hilfe einer Hauptkomponentenanalyse geschätzt. Anders als im IAB-Modell wird in der QuBe-Bevölkerungsprojektion die Zuwanderung über das Modell TINFORGE bestimmt. Dort wird für jedes Herkunftsland der Zuwandernden entschieden, ob die Auswanderung aus dem Herkunftsland aufgrund der demografischen, sozioökonomischen oder politischen Situation vor Ort motiviert ist.

Im Jahr 2019 wird die Qualifikations- und Berufsprojektionen erstmals auf Ebene von Arbeitsmarktregionen erstellt. Dazu ist es erforderlich, die QuBe-Bevölkerungsprojektion zunächst für die Kreisebene zu erstellen und diese Ergebnisse zu Arbeitsmarktregionen zu aggregieren. Da seitens des Statistische Bundesamt jedoch nur Ergebnisse auf Ebene der Bundesländer vorliegen, werden hierfür als weitere Informationsquelle die von den einzelnen Bundesländern erstellten Bevölkerungsprojektionen auf Kreisebene hinzugezogen.

Im Vortrag werden zum einen die verwendeten Modelle und Methoden der QuBe-Bevölkerungsprojektion dargestellt, das Vorgehen zur Regionalisierung der QuBe-Bevölkerungsprojektion aufgezeigt und zum anderen ein Vergleich zur 14. koordinierten Bevölkerungsvorausberechnung des Statistischen Bundesamtes auf Bundesländerebene gezogen und auf Vorzüge und Nachteile des gewählten Vorgehens zur Regionalisierung der Ergebnisse hingewiesen.

100 Ergebnisse der VGR-Revision 2019 in neuem Gewand – das VGR-Dashboard

Stefan Hauf; Tanja Mucha

Statistisches Bundesamt, Germany
tanja.mucha@destatis.de

SECTION: National Accounts, Welfare Measurement

Im August 2019 wird das Statistische Bundesamt die ersten Ergebnisse der Generalrevision der Volkswirtschaftlichen Gesamtrechnungen (VGR) 2019 vorlegen. Solche Generalrevisionen der VGR-Daten finden in Deutschland etwa alle fünf Jahre statt, um neue Definitionen, Konzepte und/oder Datenquellen in das Rechenwerk der VGR einzuführen. Wie üblich wird diese Gelegenheit genutzt, die gesamten VGR-Systeme umfassend und für einen längeren Zeitraum zu überprüfen und – wenn nötig – zu überarbeiten. Neue Erkenntnisse und Ausgangsdaten werden dieses Mal zurück bis zum Jahr 1991 in die Berechnungen integriert. Die Generalrevision 2019 ist Teil einer terminlich harmonisierten Revision in Europa, an der fast alle europäischen Mitgliedstaaten teilnehmen. Der Beitrag beleuchtet die gravierendsten Änderungen der VGR-Revision 2019 auf Niveau und Veränderungsraten des Bruttoinlandsprodukt (BIP) und der wichtigsten VGR-Aggregate.

Neu sind aber nicht nur sämtliche Ergebnisse der VGR ab 1991, sondern auch ihre interaktive graphische Darstellung auf den Internet-Seiten des Statistischen Bundesamtes: Das sogenannte VGR-Dashboard bietet den Nutzern stark nachgefragte gesamtwirtschaftliche Daten, die durch Grafiken illustriert und als interaktive Ergebnisübersichten in kompakter, attraktiver Form in das neue Web-Angebot des Statistischen Bundesamtes eingebettet sind.

Für wesentliche Wirtschaftsindikatoren aus den VGR kann mit wenigen Mausklicks über die Darstellung von Inhalten, Zeitraum und Periodizität entschieden werden. Trotz umfangreicher Variationsmöglichkeiten bleibt doch alles transparent, nutzerfreundlich und vor allem schnell im Zugriff. Die so erstellten Grafiken können in unterschiedlichen Formaten unmittelbar exportiert und von den Nutzerinnen und Nutzern weiterverarbeitet werden.

Der Vortrag wird beide Neuerungen aus den VGR, also die revidierten Ergebnisse der Generalrevision 2019 sowie das VGR-Dashboard, kombinieren und der interessierten Fachöffentlichkeit präsentieren.

101 BIP Flash $t+10$ – Eine Machbarkeitsstudie zur beschleunigten Schätzung des Bruttoinlandsprodukts

Xaver Dickopf; Christian Janz; Tanja Mucha

Statistisches Bundesamt, Deutschland
xaver.dickopf@destatis.de

SECTION: National Accounts, Welfare Measurement

Das Statistische Bundesamt veröffentlicht die ersten Ergebnisse des Bruttoinlandsprodukts (BIP) 45 Tage nach Quartalsende. Intern wird zusätzlich eine Schnellschätzung (engl. Flash Estimate) des BIP 30

Tage nach Quartalsende durchgeführt und für die Berechnung der EU-Aggregate an Eurostat geliefert. Diese Schnellschätzung basiert auf einem als 3-Säulen-Prognose bezeichneten Verfahren, das sich aus einer Expertenschätzung der verantwortlichen Fachreferate, einer ökonomischen Schätzung auf Basis von ARIMA-Modellen und einer abgestimmten Schätzung zusammensetzt. In die abgestimmte finale Schätzung gehen die Expertenschätzung und die ökonomische Schätzung, beide jeweils entstehungs- und verwendungsseitig berechnet, ein.

Im Rahmen einer Machbarkeitsstudie testet das Statistische Bundesamt derzeit eine weitere Beschleunigung der Schnellschätzung des BIP. Gegenstand der Machbarkeitsstudie ist zunächst eine rein ökonomische Schätzung des entstehungsseitigen BIP bereits zehn Tage nach Quartalsende ($t+10$).

Da zu diesem frühen Zeitpunkt kaum amtliche Basisdaten zur Verfügung stehen, werden zusätzlich auch verschiedene amtliche und nicht-amtliche Konjunkturindikatoren in die Schätzungen einbezogen. Im Rahmen des Beitrags wird daher besonders auf die Datenlage zehn Tage nach Quartalsende und die verwendeten nicht-amtlichen Datenquellen eingegangen.

Methodisch orientieren sich die Schätzungen der Machbarkeitsstudie an ökonomischen Modellen, die typischerweise im Bereich Prognose bzw. Nowcasting von makroökonomischen Zeitreihen verwendet werden. So finden bisher ARIMA-Modelle und Brückengleichungen Anwendung in den Schätzungen. Mögliche alternative Methoden wie strukturelle Zeitreihenmodelle, Faktormodelle oder Methoden aus dem Bereich Machine Learning werden im Beitrag ebenfalls erörtert.

Erste vorläufige auf Basis von historischen Echtzeitdaten geschätzte Ergebnisse des entstehungsseitigen BIP für den Testzeitraum vom ersten Quartal 2016 bis zum zweiten Quartal 2019 werden vorgestellt und Kriterien zur Beurteilung der Prognosequalität diskutiert. Den Abschluss bildet eine Revisionsanalyse der ersten Testergebnisse für das BIP nach zehn Tagen. Die abschließende Auswertung der Machbarkeitsstudie wird im Jahr 2020 erwartet.

102 Globalisierung in den VGR: Erste Ergebnisse von Pilotuntersuchungen auf europäischer Ebene

Markus Ahlborn; Thorsten Haug

Statistisches Bundesamt, Deutschland
thorsten.haug@destatis.de

SECTION: National Accounts, Welfare Measurement

Die amtliche Statistik hat den Anspruch, die notwendigen statistischen Informationen für die Willensbildung und Entscheidungsprozesse in einer demokratischen Gesellschaft zu liefern. Wichtige volkswirtschaftliche Indikatoren wie das Bruttoinlandsprodukt (BIP) bilden z.B. die Grundlage für viele geld-, fiskal- und sozialpolitische Entscheidungen. Dieser Anspruch ist allerdings gefährdet, wenn Zweifel an der konzepttreuen und kohärenten Erfassung der zugrundeliegenden Daten bestehen.

Die konzepttreue Erfassung wird derzeit in Bezug auf Multinationale Unternehmensgruppen (MUG) hinterfragt, da ihre globalen Wertschöpfungsketten in manchen Fällen nur mit großem Aufwand in statistische Erhebungen überführt werden können. Dies gilt insbesondere für das Konzept des wirtschaftlichen Eigentums, das in den Volkswirtschaftlichen Gesamtrechnungen eine zentrale Rolle spielt und damit u.a. für die Berechnung des BIP eines Landes wesentlich ist.

Spätestens mit dem „Irischen Fall“ – das irische BIP wuchs 2015 allein auf Grund von Verlagerungen geistigen Eigentums und assoziierter Waren- und Dienstleistungsexporte durch einige wenige MUG zu-

sätzlich um 20,9% – ist das Thema der Zuverlässigkeit von BIP-Berechnungen weltweit auf die Agenda der statistischen Ämter gekommen.

In einem Eurostat-Pilotprojekt zu MUG werden derzeit exemplarische Untersuchungen zu den Auswirkungen multinationaler Aktivitäten auf das Bruttonationaleinkommen in Zusammenarbeit aller nationalen Statistikämter der EU durchgeführt, die Mitte 2019 abgeschlossen sein werden. Im Beitrag werden erste Erkenntnisse dieser Pilotstudien zusammengefasst. Abschließend wird vor diesem Hintergrund kurz thematisiert, welchen Beitrag zukünftig eine Large Cases Unit zur adäquaten statistischen Erfassung von großen, komplexen oder multinationalen Unternehmensgruppen in der deutschen amtlichen Statistik leisten kann.

103 Understanding Child Deprivation in Indonesia: The Multiple Overlapping Deprivation Analysis (MODA) Approach

Ayesha Tantriana; Sri Hartini Rachmad

Statistics Indonesia, Indonesia
ayeshatantriana@gmail.com

SECTION: National Accounts, Welfare Measurement

Poverty is multidimensional and can be measured in many ways, not only by looking at the monetary aspect. Over the last two decades, multidimensional deprivation analysis has been introduced as an attempt to address the poverty issues. The Multiple Overlapping Deprivation Analysis (MODA) is a methodology developed by UNICEF which provides a comprehensive approach to the multidimensional aspects of child poverty and deprivation. It encompasses a large set of tools ranging from indicator and dimension headcounts, overlaps between several dimensions, decomposition of the adjusted multidimensional deprivation headcounts, to overlaps between monetary poverty and multidimensional deprivation. This paper demonstrates the application of the MODA methodology to children aged 0-17 years in Indonesia, using Indonesia Social and Economic Survey (SUSENAS) 2015. The result shows that in aggregate, 78 of 100 children aged 0-17 years in Indonesia are multi-dimensionally deprived. For children aged 0-4 years, most frequent deprivation happens in health dimension, while children aged 5-17 years, most frequent deprivation happens in facility dimension. In both age group, the percentage of deprived children in rural are more than the percentage of deprived children living in urban, specifically simultaneous deprived in at least two dimensions.

104 Explanation, Exploration and Debugging of Predictive Machine Learning Models

Przemyslaw Biecek

Warsaw University of Technology, Polen
przemyslaw.biecek@gmail.com

SECTION: Computational Statistics and Data Analysis

Complex predictive models are very popular and the popularity is driven by high performance in Kaggle style competitions. Methods like deep neural networks, or ensembles created with boosting, bagging or stacking, are very elastic and able to grasp non linear relations among features. But the complexity creates various risks and debts. Opacity undermines trust, lack of understanding is a problem when models are to be deployed in regulated environments like credit scoring or personalized medicine.

During the presentation I will present examples in which concept drift or lack of stability lead to unfair decisions or significant deterioration of model performance. Sometimes even below the acceptable threshold. Recent changes in law (see for example discussion about the Right to explanation in General Data Protection Regulation) increase interest in development of new methods for model explanations, exploration, auditing, debugging and stress testing. We cannot improve if we do not understand the problem.

During the talk I will present selected techniques like LIME, SHAP, Ceteris Paribus, iBreakDown, localModel that can be used for exploration, explanation and debugging of predictive models. I will discuss differences and similarities between these methods and share our experience from applications of these techniques to real large scale medical problems.

105 Rage Against The Machine – The limits of machine learning for automatic product classification

Jens Mehrhoff

European Commission, Luxembourg
Jens.MEHRHOFF@ec.europa.eu

SECTION: Computational Statistics and Data Analysis

The talk contains an assessment of applying supervised learning classification algorithms to automatically map scanner data to the official classification of sub-indices for consumer price statistics. The main test objectives are summarised as follows:

- compare the family of probabilistic supervised machine learning classification methods;
- use more or less detailed product descriptions and identifiers as feature-generating variables; and
- consider different percentages of the data set as training data – from smaller and larger data sets.

A major challenge in this respect is textual feature engineering. The product descriptions are not natural text but use specific vocabularies and rely on different kinds of shorthand. To this end, all possible three-character strings (trigrams) are exploited.

The performance measures derived are precision and recall as well as the execution time. Particular emphasis will be given to the interpretation of the results in terms of usability in the production process, i.e. guidance on what works, what does not, and why.

106 How well can Noncognitive Skills Predict Unemployment? A Machine Learning Approach

Jana Marečková (1); Winfried Pohlmeier (2)

1: Universität St. Gallen, Schweiz; 2: Universität Konstanz, Deutschland
Winfried.Pohlmeier@uni-konstanz.de

SECTION: Computational Statistics and Data Analysis

We study the predictive quality of noncognitive skill measures by means of machine learning techniques. Unlike previous empirical approaches centering around the in-sample explanatory power of non-cognitive skills, our approach focuses on the performance of predicting individual unemployment over a long-run horizon of 20 years and more. Our machine learning approach can cope not only with the challenge of selecting the most relevant factors from data with a large number of skill measures but also leads to a sparse set of skill measures which is economically and psychologically interpretable. Using data from the British Cohort Study, we compare the predictive power of different noncognitive skill measures and illustrate, how our estimates can be used to optimize the assignment mechanisms for manpower training programs and psychological intervention schemes for youths and young adults.

107 Functional Time Series Analysis

Nazarii Salish

University Carlos III de Madrid, Spain
nsalish@eco.uc3m.es

SECTION: Statistical Theory and Methods

The field of functional time series analysis has seen rapid development in the last decade. Its methods and central ideas have been broadly applied in medicine, engineering, finance, environmental studies and economics. As the literature is evolving fast it is not possible to give a full account of all current research activities within this field. However in this work we review most established concepts and some of the recent developments in FTSA covering key notions of dependence, dynamic regression model, forecasting techniques (for both linear and nonlinear processes) and tools for inference. In particular, testing procedures for stationarity and predictive accuracy are discussed. Two empirical examples are used to illustrate the performance of the discussed tools: (a) electricity consumption in nordic countries and (b) Yield curves in US.

108 Depth for curve data and applications

PierreLafaye De Micheaux (1); Pavlo Mozharovskyi (2); Myriam Vimond (3)

1: School of Mathematics and Statistics, UNSW Sydney, Australia; 2: LTCI, Telecom ParisTech, University of Paris-Saclay, France; 3: CREST, Ensaï, Bretagne Loire University, France
pavlo.mozharovskyi@telecom-paristech.fr

SECTION: Statistical Theory and Methods

John W. Tukey (1975) defined statistical data depth as a function that determines centrality of an arbitrary point with respect to a data cloud or to a probability measure. During the last decades, this seminal idea of data depth evolved into a powerful tool proving to be useful in various fields of science. Recently, extending the notion of data depth to the functional setting attracted a lot of attention among theoretical and applied statisticians. We go further and suggest a notion of data depth suitable for data represented as curves, or trajectories, which is independent of the parametrization. We show that our curve depth satisfies theoretical requirements of general depth functions that are meaningful for trajectories. We apply our methodology to diffusion tensor brain images and also to pattern recognition of hand written digits and letters.

109 Testing the predictive accuracy of functional forecasts

Matei Demetrescu (1); Nazarii Salish (2)

1: CAU Kiel, Deutschland; 2: University Carlos III de Madrid
mdeme@stat-econ.uni-kiel.de

SECTION: Statistical Theory and Methods

Models for time series where the data points are functions are often used for forecasting purposes. We compare the accuracy of competing functional forecasts as measured by the L2 norm of the functional forecast errors and derive the limiting distribution of a Diebold-Mariano type under estimation error. To this end, we consider a particular family of regression-based forecasts in detail, and then provide generic conditions allowing one to test the equal predictive accuracy of various functional prediction methods.

110 Längsschnittverknüpfung von MZ-Panel-Daten

Hans-Ullrich Mühlenfeld (1); Sven Brocker (2)

1: IT.NRW - Statistisches Landesamt NRW, Deutschland; 2: Universität Duisburg
hans-ullrich.muehlenfeld@it.nrw.de

SECTION: FDZ

Seit 2012 besteht die Möglichkeit, mehrere Mikrozensusjahrgänge miteinander zu verknüpfen. Mittels zeitstabiler Identifikatoren kann aus der grundsätzlich als Querschnittserhebung angelegten Erhebung ein Längsschnittdatensatz erstellt werden. Der Vortrag zeigt, welche Herausforderungen es bei der Verknüpfung der Jahrgänge 2012 bis 2015 gibt, wie das konkrete Vorgehen aussieht und wofür ein solcher Paneldatensatz letztlich genutzt werden kann. Darüber hinaus werden Herausforderungen angesprochen, für die es so noch keine eindeutige Lösung gibt.

111 Entwicklung Synthetischer Datenstrukturfiles in den FDZ

Janina Loske (1); Tobias Wolfanger (2)

1: Statistisches Bundesamt, Deutschland; 2: IT.NRW
janina.loske@destatis.de

SECTION: FDZ

Die Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder (FDZ) stellen für wissenschaftliche Projekte im Rahmen des Datenzugangs per kontrollierter Datenfernverarbeitung (KDFV) sogenannte Datenstrukturfiles zu Verfügung, um anhand dieser die Entwicklung von Auswertungsprogrammen zu ermöglichen. Bei den Datenstrukturfiles handelt es sich um Mikrodatsätze, die ein technisch vergleichbares, absolut anonymes, verkleinertes Abbild des im FDZ für die Auswertung per KDFV bereitliegenden Datenmaterials darstellen. Durch die bisher gewählten Maßnahmen zur Umsetzung der gesetzlich erforderlichen Anonymisierung des Datenmaterials gehen jedoch auch die inhaltlichen Abhängigkeitsstrukturen zwischen den Merkmalen verloren.

Für den Mikrozensus und die Fallpauschalenbezogene Krankenhausstatistik (DRG-Statistik) entwickeln die FDZ zurzeit synthetische Datenstrukturfiles. Synthetische Datenstrukturfiles sollen neben der technischen Struktur der Daten auch Abhängigkeitsstrukturen und multivariate Verteilungen der Originaldaten abbilden, ohne von der Grundvoraussetzung der absoluten Anonymität abzuweichen. Für die Erstellung dieser synthetischen Datenstrukturfiles werden im Sinne des maschinellen Lernens Eigenschaften eines Originaldatenmaterials erlernt und dann durch Vorhersage mittels des erlernten Modells auf eine synthetische Population übertragen. Bei den Modellen kann es sich z.B. um multinomiale logistische Regressionsmodelle oder Entscheidungsbäume handeln. Die Entwicklung der synthetischen Datenstrukturfiles erfolgt mit Hilfe der R-Pakete „simPop“ für den Mikrozensus und „synthpop“ für die DRG-Statistik.

Durch die Bereitstellung synthetischer Datenstrukturfiles im Rahmen der KDFV wird angestrebt, der Wissenschaft nicht nur die Möglichkeit einer syntaktischen Prüfung ihrer Auswertungsprogramme zu geben, sondern auch Einsichten zu multivariaten Verteilungen und Abhängigkeitsstrukturen in Anlehnung

an das Originaldatenmaterial zu ermöglichen um damit ein zielorientierteres inhaltliches Arbeiten zu erleichtern.

112 Der EU-Unternehmensbegriff – was heißt das für die Firmendaten der amtlichen Statistik in Deutschland (AFiD-Daten) im FDZ

Ramona Voshage (1); Michael Rößner (2)

1: Amt für Statistik Berlin-Brandenburg, Deutschland; 2: Statistisches Landesamt Sachsen-Anhalt
ramona.voshage@statistik-bbb.de

SECTION: FDZ

Die adäquate Definition und Abgrenzung einer statistischen Einheit ist wesentlich für jede Datengrundlage. In den amtlichen Firmendaten für Deutschland (AFiD) ist das „Unternehmen“ derzeit die zentrale Analyseeinheit. Die deutsche amtliche Statistik definiert aktuell ein Unternehmen als kleinste rechtliche Einheit. Da jedoch die amtlichen Unternehmensstatistiken zukünftig komplexe Produktionsstrukturen und Unternehmensverflechtungen besser abbilden sollen, fordern europäische Datenanforderungen einen erweiterten Unternehmensbegriff anzuwenden.

Die EU-Einheitenverordnung gibt eine weitergehende Definition vor: Das Unternehmen entspricht der kleinsten Kombination rechtlicher Einheiten, die eine organisatorische Einheit zur Erzeugung von Waren und Dienstleistungen bildet und insbesondere in Bezug auf die Verwendung der ihr zufließenden laufenden Mittel über eine gewisse Entscheidungsfreiheit verfügt. Ein Unternehmen übt eine Tätigkeit oder mehrere Tätigkeiten an einem Standort oder an mehreren Standorten aus. Ein Unternehmen kann einer einzigen rechtlichen Einheit entsprechen.

Für die Statistik bedeutet dies also, dass der Zusammenhang und die Kontrollstrukturen rechtlicher Einheiten geprüft und sinnvolle Zusammenfassungen vorgenommen werden müssen (Profiling). Die bisherigen Erhebungen sollen dabei möglichst unangetastet werden, so dass die wie bisher erhobenen Ergebnisse auf Ebene der rechtlichen Einheiten zu Angaben der Unternehmen zusammengeführt werden müssen (Imputation und Konsolidierung).

Auf die AFiD-Produkte hat das Ganze somit zunächst keinen direkten Einfluss. Es ist natürlich zu prüfen, inwiefern imputierte Einheiten Eingang in die FDZ-Daten finden und ob ein Unternehmensidentifikator den Zusammenhang der bisher enthaltenen Angaben der rechtlichen Einheiten abbilden wird.

Der Beitrag beschreibt, welche Auswirkungen die Umstellung des EU-Unternehmensbegriffs auf die Mikrodaten der amtlichen Firmendaten für Deutschland haben wird.

113 Was kann die Statistik der Bundesagentur für Arbeit (BA) zum Thema Wohnen aussagen? Wie wohnen SGB-II-Leistungsberechtigte?

Claudia Korn

Bundesagentur für Arbeit, Germany
claudia.korn@arbeitsagentur.de

SECTION: Housing

Hört man von der Statistik der BA, denkt man oft zu aller erst an Arbeitslosenzahlen und möglicherweise noch an die Beschäftigtenstatistik. Dabei gibt es noch viele andere Themen, zu denen die BA Daten bereitstellt – darunter auch die Wohnsituationsstatistik.

Der Vortrag gibt einen Einblick in die Wohnkosten- und Wohnsituationsstatistik, die die Statistik der Bundesagentur für Arbeit im Bereich der Grundsicherung für Arbeitsuchende (Zweites Sozialgesetzbuch SGB II, umgangssprachlich: „Hartz IV“) erstellt. Außerdem werden Besonderheiten und Vorteile aber auch ihre Grenzen vorgestellt. Abschließend zeigen Anwendungsbeispiele, wofür man diese Daten nutzen kann.

Mit der Grundsicherungsstatistik SGB II können Fragen zur Wohnsituation von SGB-II-Leistungsberechtigten beantwortet werden. Typische Merkmale sind Unterkunftsort (Miete, Eigentum, Tagessatz), Wohnfläche sowie Wohnkosten. Als Sekundärstatistik nutzt sie die Verwaltungsdaten aller Jobcenter, die für die Leistungsgewährung zuständig sind. Die Daten umfassen die Unterkunftsinformationen von ca. 3 Mio. Bedarfsgemeinschaften (Haushalte), in denen ca. 5,4 Mio. Leistungsberechtigte leben. Die Wohnsituationsstatistik wird monatlich für Deutschland und die Bundesländer sowie für Kreise und Jobcenter veröffentlicht.

Um die Vergleichbarkeit bzw. Vereinbarkeit zu anderen Statistiken, die sich mit dem Thema Wohnen beschäftigen, analysieren zu können, werden die Ergebnisse der Wohnsituationsstatistik der BA beispielsweise in Beziehung gesetzt zu Daten des Zensus 2011.

Anhand von Anwendungsbeispielen werden die Möglichkeiten der Statistik dargestellt. Wie haben sich beispielsweise die Wohnkosten für die SGB-II-Leistungsberechtigten bis 2018 entwickelt?

114 Wie wohnen geflüchtete Menschen in Deutschland, die auf Leistungen der Grundsicherung (SGB II) angewiesen sind? – Eine Analyse struktureller und regionaler Besonderheiten.

Christiane Tentscher

Statistik der Bundesagentur für Arbeit, Deutschland
christiane.tentscher@arbeitsagentur.de

SECTION: Housing

In diesem Vortrag wird der Frage nachgegangen, wie geflüchtete Menschen – die auf Leistungen der Grundsicherung nach dem zweiten Sozialgesetzbuch angewiesen sind – in Deutschland wohnen.

Die Statistik der Bundesagentur für Arbeit bietet als Sekundärstatistik die Möglichkeit auf eine Vollerhebung zum Thema „Wohnen im Leistungsbezug SGB II“ zurückzugreifen. Die Verwaltungsdaten aller Grundsicherungsträger werden dafür monatlich aufbereitet und veröffentlicht – insgesamt liegen damit Informationen zu etwa 3 Millionen Haushalten und gut 5 Millionen Menschen in der Grundsicherung (SGB II) vor. Darunter auch etwa 400.000 Haushalte (Bedarfsgemeinschaften) in denen Geflüchtete leben.

Der Vortrag soll der Frage auf den Grund gehen, inwiefern sich die Bedarfsgemeinschaften in denen geflüchtete Menschen leben von denen ohne Migrationserfahrung unterscheiden. Gibt es strukturelle Besonderheiten? Welche Unterschiede hinsichtlich des Wohnens zeigen sich? Welche zusätzlichen Anforderungen an die Wohnungsmärkte entstehen, wenn sich die wohnungsbezogenen Unterschiede auflösen? Und welche Rolle spielt die Region (v.a. unter siedlungsstrukturellen Aspekten), in denen geflüchteten Menschen unterkommen?

Antworten auf diese und weitere Fragen, werden unter anderem über Kennzahlen aus der Statistik der Grundsicherung für Arbeitssuchende sowie aus Verknüpfungen zu anderen Statistiken der Bundesagentur für Arbeit analysiert und grafisch vorgestellt.

115 Sinus-Geo-Milieus[®] Migranten

Guido Oemmelen

microm, Germany
g.oemmelen@microm.de

SECTION: Housing

Menschen mit Migrationshintergrund stellen einen bedeuteten Anteil an der Bevölkerung Deutschlands dar. Die Definition von Menschen mit Migrationshintergrund basiert hier auf der Definition des Statistischen Bundesamtes. Von aktuell 41 Mio. Haushalten in Deutschland haben dementsprechend ca. 23,46% einen Migrationshintergrund.

Die Zugehörigen dieser großen Bevölkerungsgruppe unterscheiden sich hinsichtlich Grundorientierung, Werten bzw. Lebensstilen teilweise immens. Die Milieu-Landkarte zeigt für jedes Milieu ein modellhaftes Abbild der sozialen Schicht und der Wertestruktur. Die Forschung zeigt, „DIE“ Migranten gibt es nicht. Eine aktuelle Studie vom vhw (Bundesverband für Wohnen und Stadtentwicklung) beschreibt, dass es in der Migrantenpopulation, ebenso wie in der autochthonen Bevölkerung, eine große Vielfalt von Lebensauffassungen und Lebensweisen gibt. Identifiziert werden 10 Migranten-Milieus mit jeweils unterschiedlichen Lebenswelten, Wertebildern und Integrationsniveaus.

Residentielle Segregation beschreibt das Phänomen, dass einzelne Bevölkerungsgruppen sich in bestimmten Stadtteilen oder Vierteln konzentrieren und somit räumlich getrennt leben von anderen Bevölkerungsgruppen. Üblicherweise wird Segregation als Folge von Diskriminierung oder mangelnden Integrationswillen gedeutet. Dies muss jedoch nicht zwangsläufig der Fall sein. Wo Menschen leben und wohnen, muss nicht unbedingt auf ihren Status als Zuwanderer zurückgehen, es kann sich auch aus sozio-ökonomischen Eigenschaften wie dem Bildungsabschluss und dem Einkommen ergeben. Die Segregation zwischen Statusgruppen könnte zur Segregation von Migrantengruppen führen, wenn ausgeprägte Statusunterschiede zwischen Einheimischen und Migranten und zwischen verschiedenen Gruppen von Migranten bestehen.

Der Vortrag stellt die Vorgehensweise der kleinräumigen Verortung der Migrantenmilieus über Small Area Statistics Verfahren in den Fokus sowie die Vorstellung und Diskussion ausgewählter Ergebnisse anhand kartografischer Darstellungen.

116 Unbesetzte Ausbildungsplätze – Welche regionalen und betrieblichen Gründe spielen eine Rolle?

Sandra Dummert; Ute Leber; Barbara Schwengler

Institut für Arbeitsmarkt- und Berufsforschung, Deutschland
Barbara.Schwengler@iab.de

SECTION: Labour Market and Social Security

Die aktuelle Situation auf dem deutschen Lehrstellenmarkt ist von zwei gegenläufigen Entwicklungen geprägt. Auf der einen Seite haben die Betriebe immer mehr Probleme bei der Besetzung ihrer Ausbildungsplätze und die Zahl der freien Ausbildungsplätze steigt. Andererseits ist der Übergang in die Ausbildung für viele Jugendliche nach wie vor schwierig und die Zahl der erfolglosen Bewerber steigt. Die Abstimmung von Angebot und Nachfrage wird daher auf dem deutschen Stellenmarkt für Ausbildungsplätze immer schwieriger. Mit dem Linked Employer-Employee Datensatz (LIAB) des Instituts für Arbeitsmarkt- und Berufsforschung (IAB) können wir das Phänomen der unbesetzten Ausbildungsplätze genauer untersuchen. Die Daten erlauben uns einen Überblick über die Entwicklung der offenen Ausbildungsplätze in Ost- und Westdeutschland. Zudem können wir Faktoren identifizieren, die für die Nichtbesetzung der Ausbildungsplätze verantwortlich sind. Wir gehen davon aus, dass die offenen Stellen nicht nur auf interne betriebliche Gründe, wie die Betriebsgröße oder das Lohnangebot für Auszubildende, sondern auch auf externe Bedingungen, wie allgemeine regionale Faktoren, zurückzuführen sind. Daher betrachten wir auch die Situation auf der Nachfrageseite des Arbeitsmarktes innerhalb einer Region. Die Ergebnisse unserer mehrstufigen Mixed-Effects-Schätzungen zeigen, dass neben den Merkmalen auf Unternehmensebene auch regionale Determinanten den Anteil der freien Ausbildungsplätze beeinflussen. So ist für den Rekrutierungserfolg von Unternehmen nicht nur die Attraktivität des Unternehmens von Bedeutung, sondern auch die Verfügbarkeit von Schulabgängern in einer Region und der regionale sektorale Wettbewerb. Unsere Analyse zeigt auch erhebliche Unterschiede zwischen Ost- und Westdeutschland. Dies betrifft sowohl die Zahl der unbesetzten Ausbildungsplätze als auch die sie beeinflussenden Faktoren.

117 Berufliche Mobilität von Arbeitslosen

Michael Hartmann

Bundesagentur für Arbeits, Deutschland
Michael.Hartmann@arbeitsagentur.de

SECTION: Labour Market and Social Security

Berufliche Mobilität ist ein wichtiger Bestandteil der Dynamik am Arbeitsmarkt. Digitalisierung und demografischer Wandel werden den beruflichen Strukturwandel beschleunigen und den Arbeitskräften noch mehr berufliche Mobilität abverlangen. Die Anforderungen an das statistische Berichtssystem für solche Fragestellungen belastbare Informationen zur Verfügung zu stellen nehmen zu. Der Vortrag befasst sich mit der beruflichen Mobilität von Arbeitslosen. Wie kann berufliche Mobilität in der Arbeitslosenstatistik der Bundesagentur für Arbeit gemessen werden? In welchem Ausmaß kann bei Arbeitslosen

berufliche Mobilität beobachtet werden und welche Aussagekraft haben die Mobilitätsergebnisse? In der Arbeitslosenstatistik der Bundesagentur für Arbeit können die beruflichen Informationen über Arbeitslose in Informationen über den Ausbildungs-, den Herkunfts-, den Ziel- und den Einmündungsberuf unterschieden werden. Berufliche Mobilitätsprozesse lassen sich über einen paarweisen Vergleich von Ausbildungs-, Herkunfts-, Ziel- und Einmündungsberuf feststellen. Der Vortrag erläutert die methodischen Überlegungen zur Messung der beruflichen Mobilität und stellt ausgewählte Ergebnisse vor. ,

118 Beschäftigungsstabilität nach Langzeitarbeitslosigkeit

Thomas Rothe

Institut für Arbeitsmarkt- und Berufsforschung, Deutschland
thomas.rothe@iab.de

SECTION: Labour Market and Social Security

Hohe und persistente Langzeitarbeitslosigkeit bleibt weiter ein zentrales Problem am deutschen Arbeitsmarkt. Dies gilt ungeachtet der zuletzt sinkenden Zahl an Langzeitarbeitslosen mit einer Arbeitslosigkeitsdauer von mindestens einem Jahr. Trotz verschiedener Schwierigkeiten gibt es eine bestimmte Gruppe von vormals Langzeitarbeitslosen, denen der Weg zurück in Beschäftigung gelingt. Wir untersuchen die Einmündung Arbeitsloser in eine möglichst stabile und ungeförderte Beschäftigung, sei es unmittelbar nach der Arbeitslosigkeit oder nach diversen anderen Erwerbs- oder Nichterwerbsphasen. Die Datengrundlage bildet eine zweiprozentige Zufallsstichprobe der Integrierten Erwerbsbiographien (IEB) des Instituts für Arbeitsmarkt- und Berufsforschung der Bundesagentur für Arbeit (BA).

Erste Analysen zeigen, dass Beschäftigungsverhältnisse unmittelbar nach Langzeitarbeitslosigkeit im Durchschnitt etwas länger sind als Beschäftigungen nach Kurzeitarbeitslosigkeit. Beschäftigungsverhältnisse nach Langzeitarbeitslosigkeit, die nach den Hartz-Reformen im Jahr 2005 und insbesondere nach der großen Rezession 2009 begonnen wurden, scheinen deutlich stabiler zu sein als in den Jahren 2002 und 2003. Für die längeren Beschäftigungsdauern nach einer Langzeitarbeitslosigkeit dürften einerseits Selektionseffekte eine Rolle spielen, d.h. dass Personen nach längerer Arbeitslosigkeit insbesondere dann eingestellt werden, wenn sie vergleichsweise gut qualifiziert sind oder aufgrund anderer mit unseren Daten nicht beobachtbarer Merkmale eine relative hohe Produktivität erwarten lassen. Zum anderen dürften Beschäftigte nach einer längeren Arbeitslosigkeitsphase besonders darauf bedacht sein ihre Beschäftigung nicht zu verlieren oder freiwillig zu beenden, um das Risiko einer erneuten Arbeitslosigkeit möglichst gering zu halten.

Die Chancen aus Langzeitarbeitslosigkeit heraus eine stabile, ungeförderte, sozialversicherungspflichtige Beschäftigung von mindestens einem Jahr zu beginnen, lagen pro Monat bei durchschnittlich 0,5 Prozent (2,9 Prozent für Kurzeitarbeitslose). Im Vergleich zu den direkten Übergängen nach Arbeitslosigkeit, werden fast fünfmal so viele stabile Beschäftigungsverhältnisse innerhalb von fünf Jahren nach einer Langzeitarbeitslosigkeit begonnen. Langzeitarbeitslose Personen, die innerhalb von fünf Jahren nach Ende der Arbeitslosigkeit in eine stabile Beschäftigung einmünden, unterscheiden sich hinsichtlich Alter, Berufsausbildung und Geschlecht von anderen Langzeitarbeitslosen.

119 Mögliche Auswirkungen des Brexit auf die künftige Zuwanderung nach Deutschland und die Beschäftigungsentwicklung einheimischer Betriebe

Mario Bossler (1); Johann Fuchs (1); **Alexander Kubis (1)**; Lutz Schneider (2)

1: Institut für Arbeitsmarkt- und Berufsforschung (IAB), Deutschland; 2: Hochschule für angewandte Wissenschaften Coburg, Coburg
alexander.kubis@iab.de

SECTION: Labour Market and Social Security

Zwei Aspekte sind im Rahmen der Diskussion über die Folgen des Brexit bisher unzureichend untersucht. Zum einen stellt sich die Frage ob und wie sich die künftige Zuwanderung nach Deutschland durch den Brexit verändert, da der institutionelle Schock durch den Brexit die gesamte EU-Binnenmigration beeinflussen kann. Zum anderen kann der Brexit auch die bestehenden wirtschaftlichen Beziehungen tangieren, indem sich die (künftige) Personalnachfrage der deutschen Betriebe anpasst. Beide Punkte werden im Vortrag beleuchtet.

Als Folge des Brexit wird vermutet, dass institutionell bedingt die Migration ins Vereinigte Königreich abnimmt. Weitestgehend unbekannt sind die damit verbundenen Folgewirkungen auf die Migration in andere EU-Länder, wie beispielsweise Deutschland. Der im Beitrag verfolgte Ansatz fixiert die Verteilung der sich im Ausland befindlichen Nationalitäten innerhalb Europas und erklärt künftige Migrationsströme (beispielsweise ausgelöst durch den Brexit) mit der bestehenden Verteilung von Migranten innerhalb der EU. Ein hoher Anteil an Migranten einer bestimmten Nationalität (Diaspora) in einem bestimmten Land, wie bspw. Deutschland, erhöht die Zuwanderungswahrscheinlichkeit weiterer Angehöriger dieser Diaspora.

In einem zweiten Schritt behandelt der Vortrag die Frage, inwieweit die Betriebe in Deutschland durch den Brexit beeinflusst werden. Auf Basis der repräsentativen IAB-Stellenerhebung werden verschiedene wirtschaftliche Verflechtungen im Rahmen des Vortrags untersucht. Hierzu wurden die Betriebe im Rahmen einer Sonderbefragung im zweiten Quartal 2019 befragt. In Deutschland ansässige Betriebe können durch ihre Eigentums- und Wettbewerbsverhältnisse vom Brexit betroffen sein, aber auch über Exportbeziehungen oder die direkte Beschäftigung britischer Staatsbürger. Im Beitrag wird untersucht, ob und inwieweit sich der Brexit schon heute auf die Beschäftigungsentwicklung der vom Brexit betroffenen Betriebe ausgewirkt hat. Hierbei wird neben der Beschäftigungswahrscheinlichkeit ausländischer Staatsbürger, die (erwartete) Beschäftigungsentwicklung der Betriebe insgesamt betrachtet.

120 Wearable-based Parkinson's Disease Severity Monitoring using Deep Learning

Jann Goschenhofer (1); Franz Pfister (2); Kamer Yuksel (2); Urban Fietzek (3); Bernd Bischl (1)

1: LMU Munich, Deutschland; 2: ConnectedLife GmbH; 3: Schoen Clinic Schwabing
jann.goschenhofer@stat.uni-muenchen.de

SECTION: Computational Statistics and Data Analysis

One major challenge in the medication of Parkinson's disease is that the severity of the disease, reflected in the patients' motor state, cannot be measured using accessible biomarkers. Therefore, we develop and examine a variety of statistical models to detect the motor state of such patients based on sensor data from a wearable device. We find that deep learning models consistently outperform a classical machine learning model applied on hand-crafted features in this time series classification task. Furthermore, our results suggest that treating this problem as a regression instead of an ordinal regression or a classification task is most appropriate. For consistent model evaluation and training, we adopt the leave-one-subject-out validation scheme to the training of deep learning models. We also employ a class-weighting scheme to successfully mitigate the problem of high multi-class imbalances in this domain. In addition, we propose a customized performance measure that reflects the requirements of the involved medical staff on the model. We also propose a transfer learning technique which helps to improve model performance substantially. Our results suggest that deep learning techniques offer a high potential to autonomously.

121 Better time series forecasting using expert knowledge: The Maximum Entropy Approach

Dominik Ballreich

codecentric AG, Germany
dominik.ballreich@gmail.com

SECTION: Computational Statistics and Data Analysis

Machine learning algorithms for time series forecasting have become increasingly powerful in recent decades. Nevertheless, not only the quality of the forecasts is important, but also their acceptance by the staff. Especially with regard to automatic forecasts, distrust may arise among dispatchers. Furthermore, long-standing employees often have a detailed overview of customer behavior, market situation and other important factors. Therefore, it makes sense to include this expert knowledge in the predictions of complex algorithms. This can be achieved through the Maximum Entropy approach, which is discussed in this presentation. The approach is derived in detail and applied to real data.

122 Detecting changes in the dependence structure of a time series

Alexander Dürre; Roland Fried

Technische Universität Dortmund, Germany
alexander.duerre@udo.edu

SECTION: Computational Statistics and Data Analysis

We propose a new robust test to detect changes in the dependence structure of a time series. The test is based on empirical autocovariances of a robust transformation of the original time series. Because of the transformation we do not require any finite moments of the original time series making the test especially suitable for heavy tailed time series. We furthermore propose a lag weighting scheme which puts emphasis on changes of the autocorrelation at smaller lags. Our approach is compared to existing ones in some simulations.

123 Continuous-Discrete Filtering using the Zakai Equation: Smooth Likelihood Surface

Hermann Singer

FernUniversität in Hagen, Germany
hermann.singer@fernuni-hagen.de

SECTION: Computational Statistics and Data Analysis

The continuous time state space model consists of a dynamical equation (stochastic differential equation SDE) for the unobserved state and a measurement equation, also described by an SDE. The unnormalized state probability density given the measurements up to present time t (filter density), fulfils a stochastic partial differential equation, the Zakai equation. It is a Fokker-Planck equation with potential term containing the measurements. The normalizing constants determine the likelihood function of the system. We use time dependent measurement error covariance matrices in order to accomodate sampled data (continuous-discrete filtering).

In this talk, the Zakai equation is transformed to a backward Kolmogorov equation which is solved by Monte Carlo integration using importance sampling. As is well known, standard particle filter algorithms suffer from the problem (due to resampling), that the likelihood is not a smooth function of the model parameters. Since in the proposed algorithm no resampling is required, we obtain a smooth likelihood surface. A drawback is the large number of grid points where the filter density must be evaluated. The results are compared with approximations based on a matrix representation of the Fokker-Planck operator and particle filtering.

124 Evaluating the Rationality of Mode Forecasts

Timo Dimitriadis (1,2); Andrew Patton (3); Patrick Schmidt (2,4)

1: Universität Hohenheim; 2: HITS Heidelberg; 3: Duke University; 4: Goethe Universität Frankfurt
timo.dimi@googlemail.com

SECTION: Statistical Theory and Methods

This paper considers tests of the optimality of forecasts of the mode of some target variable. Unlike other measures of central tendency, such as the mean and median, the mode is not generally an „elicitable“ functional. However, we show that it is asymptotically elicitable, which means that there exists a sequence of elicitable functionals which converges to the mode. Building on this new result, we construct an asymptotically valid test for the rationality of mode forecasts.

We further recover the functional of central tendency which is the most appropriate to the given forecasts by setting up a GMM framework. The parameter of interest which distinguishes between the mean, median and mode is potentially unidentified for symmetric distributions. Thus, we apply GMM under weak identification in order to construct asymptotically valid confidence sets.

We apply our new tests to survey forecasts of some key macroeconomic variables and find that when interpreted as mean or median forecasts we are able to reject rationality, however when interpreted as mode forecasts we cannot. Rather than interpreting a rejection of mean forecast rationality as a motivation to consider asymmetric loss functions, we argue that survey respondents may instead be reporting a common alternative measure of central tendency, namely the mode.

125 Local Whittle Wavelet Estimation for Locally Stationary Processes

Simon Wingert; Philipp Sibbertsen

Leibniz Universität Hannover, Germany
wingert@statistik.uni-hannover.de

SECTION: Statistical Theory and Methods

In this paper we introduce an estimator for time-varying long memory of locally stationary processes. The estimator is semi-parametric and wavelet-based in the sense that it optimizes a contrast function involving the discrete wavelet transform of the process for estimating the parameter. We analyze the asymptotic properties of the estimator and establish consistency and its asymptotic normality. By a Monte Carlo simulation we study the finite sample properties of the estimator. In addition we apply the estimator to realized volatility of national stock indices with data spanning 15 years.

126 Fractional cointegration in EMU government bond markets

Michelle Voges; Philipp Sibbertsen

Leibniz Universität Hannover, Germany
voges@statistik.uni-hannover.de

SECTION: Statistical Theory and Methods

It is commonly found that the markets for long-term government bonds of EMU countries were highly integrated prior to the subprime mortgage and EMU debt crisis. In contrast to this, we show that there were periods of integration and disintegration that coincide with bull- and bear-market periods in the stock market. This finding is based on the interrelation between market integration and fractional cointegration in the context of the common currency area and confirmed by a wide array of suitable semiparametric tests for fractional cointegration. A simple econometric argument about the spectral behavior of long-memory time series leads to the conclusion that there is a stronger differentiation between bonds with different default risks during periods of disintegration, so that the dynamics of the yields implied the possibility of macroeconomic and fiscal divergence between the EMU countries long before the crisis periods.

127 Fractional trends in unobserved components models

Tobias Hartl (1,2); Rolf Tschernig (1); Enzo Weber (1,2)

1: University of Regensburg, Germany; 2: Institute for Employment Research (IAB) Nuremberg, Germany
tobias1.hartl@ur.de

SECTION: Statistical Theory and Methods

We develop a generalization of permanent-transitory decompositions that avoids prior assumptions about the long-run dynamic characteristics by modelling the permanent component as a fractionally integrated process and incorporating a fractional lag operator into the autoregressive polynomial of the cyclical component.

The model neither requires stationarity nor orthogonal permanent and transitory shocks and can be cast in state-space form. In a multivariate setup, fractional trends may exhibit different integration orders, but depend on the same stochastic shocks, leading to a cointegrated system with different integration orders. We show that our fractional UC model is able to estimate a smooth trend together with a cycle hitting all NBER recessions for US real output. In a multivariate setup, we provide evidence that income and inflation are driven by the same long-run shocks, although their trend components are of different persistence.

128 Überlegungen zur Stichprobenmethodik der neuen digitalen Verdiensterhebung

Thomas Zimmermann; Ralf Droßard; Simone Scharfe

Destatis, Deutschland
thomas.zimmermann@destatis.de

SECTION: Methodology of Statistical Surveys

Im September 2018 hat das Bundesministerium für Wirtschaft und Energie Destatis mit der Erstellung eines Konzepts zur Verzahnung der vierteljährlichen Verdiensterhebung und der alle vier Jahre durchgeführten Verdienststrukturerhebung beauftragt. Beide Erhebungen sollen ab 2021 in einer neuen digitalen Verdiensterhebung (NVE) verschmelzen. Die Zielsetzung der NVE ist die ressourcenschonende Erfüllung neuer Datenwünsche der Nutzer der Verdienststatistiken.

Eine Besonderheit besteht dabei darin, dass die zukünftige NVE sowohl zur Produktion von Struktur- wie auch Konjunkturstatistiken genutzt werden soll. Dies ist aus methodischer Sicht herausfordernd, da ein für die Schätzung von strukturellen Größen optimierter Stichprobenplan in aller Regel nicht gleichzeitig auch optimal für die Schätzung von Veränderungsraten ist und umgekehrt. Zum Umgang mit diesem Zielkonflikt können konjunktur- und strukturstatistische Belange bei der Formulierung des Optimierungsproblems für die Stichprobenplanung explizit berücksichtigt werden.

Im vorliegenden Beitrag präsentieren wir den Sachstand der Arbeiten zur Stichprobenplanung für die NVE.

129 Varianzschätzung für den EU-Indikator AROPE

Stefan Zins (1); Birgit Horneffer (2)

1: Institut für Arbeitsmarkt- und Berufsforschung, Deutschland; 2: Statistisches Bundesamt, Deutschland
birgit.horneffer@destatis.de

SECTION: Methodology of Statistical Surveys

Indikatoren zur Messung von Armut und sozialer Eingliederung stehen im Fokus der EU Strategie und werden in der Statistik über Einkommen und Lebensbedingungen von Privathaushalten in Europa (EU-SILC) gemessen. Sie bilden eine wichtige Grundlage für die Europäische Sozialstatistik.

Der sogenannte AROPE-Indikator ist dabei der Hauptindikator zur Überwachung des Zielwerts zur Armutsreduktion der EU 2020 Strategie.

Die AROPE-Rate beschreibt den Anteil derer in der gesamten Bevölkerung, der von Armut oder sozialer Ausgrenzung gefährdet ist.

Erstmals wird eine Linearisierung des AROPE Indikators vorgestellt, die zur Varianzschätzung mittels Standardverfahren verwendet und somit einfacher bei komplexen Stichprobendesign angewendet werden kann, als dies mit Resamplingverfahren möglich ist.

Es werden die Ergebnisse der AROPE-Rate Punkt- und Varianzschätzung basierend auf EU-SILC Jahresdaten aus Deutschland vorgestellt.

Zudem wird die linearisierte Varianzschätzung mit einem Bootstrap-Schätzer, einen Standardverfahren zur Varianzschätzung von nicht-linearen Statistiken, verglichen.

130 Unterjährige Schätzung der Arbeitskräfteerhebung im neuen System der Haushaltsstatistiken

Anna-Lena Wölwer; Ralf Münnich

Universität Trier, Deutschland
woelwer@uni-trier.de

SECTION: Methodology of Statistical Surveys

Das neue System der Haushaltsstatistiken wird als rotierende Panel Stichprobe durchgeführt. Die Rotation der Befragten richtet sich nach ihrer Zugehörigkeit zu den Modulen LFS, SILC, IKT und Kern. Das LFS-Modul rotiert erstmals unterjährig mit zeitlicher Gleichverteilung der Befragungen über die Wochen eines Jahres. Damit stehen zu verschiedenen Zeitpunkten für einen Teil der Befragten unterjährige Informationen aus vorherigen Befragungen zur Verfügung. Diese unterjährigen Informationen aus vorherigen Befragungen können zur unterjährigen Schätzung herangezogen werden. In der Literatur werden verschiedene Schätzmethode diskutiert, mit denen die Informationen aus vorherigen Befragungen in die Schätzung einbezogen werden können. Für die Wahl einer Schätzmethode sind die Überlappungen zwischen verschiedenen Zeitpunkten sowie die zeitliche Korrelation der interessierenden Variablen relevant.

Da die unterjährige Rotation im LFS neu eingeführt wird, sind zur Evaluation der verschiedenen Schätzmethode keine unterjährigen Daten verfügbar. Es steht jedoch eine Stichprobe der Integrierten Arbeitsmarktbiografien (SIAB) 1975-2014 zur Verfügung. Die SIAB Daten umfassen die Erwerbsverläufe von 1.707.228 Personen in 51.987.959 Datenzeilen. Sie werden in Hinblick auf zeitliche Korrelationen und saisonale Strukturen analysiert. Basierend auf diesen Erkenntnissen werden Schlüsse für die unterjährige Schätzung von Erwerbstätigkeit im neuen System der Haushaltsstatistiken gezogen.

131 Ein multivariater GREG Schätzer zur Schätzung von Veränderungen über die Zeit in Rotationsstichproben

Anne Konrad (1); Yves Berger (2)

1: Universität Trier, Deutschland; 2: University of Southampton, England
konrada@uni-trier.de

SECTION: Methodology of Statistical Surveys

Die Schätzung von Veränderungen über die Zeit ist von großem Interesse sowohl für Wissenschaftler verschiedener Fachrichtungen als auch für politische Entscheidungsträger. Für die Schätzung von Veränderungen müssen die interessierenden Variablen für mindestens zwei verschiedene Zeitpunkte beobachtet werden. Panelsurveys werden häufig als Rotationsstichprobe realisiert, wobei die Beobachtungen für eine vordefinierte Anzahl an Perioden in der Stichprobe verbleiben und anschließend wieder herausrotieren. Die Herausforderung des Rotationsdesigns für die Schätzung stellt die teilweise Überlappung der Stichproben zu den verschiedenen Zeitpunkten dar. So liegen für neu hereinrotierende Einheiten keine Beobachtungen

aus der Vorperiode vor. Für herausrotierende Einheiten hingegen liegen keine Beobachtungen für die nachfolgenden Zeitpunkte vor. Um dennoch alle zur Verfügung stehenden Informationen zu nutzen, schlägt dieser Beitrag einen multivariaten GREG Schätzer vor. Dieser kombiniert die einzelnen Stichproben beider Zeitpunkte zu einer gemeinsamen Stichprobe, einschließlich der Beobachtungen, die für nur einen Zeitpunkt vorliegen. Durch die Einbeziehung eines weiteren Hilfsmerkmals, welches das Rotationsdesign beschreibt, erfasst der vorgeschlagene multivariate GREG Schätzer zusätzlich die Stichprobenkorrelation zwischen den Beobachtungen, die durch die Überlappung der Stichproben entsteht. Damit berücksichtigt der multivariate GREG Schätzer alle zur Verfügung stehenden Informationen über das interessierende Merkmal und die Hilfsmerkmale sowie die Designinformation gegeben durch das Rotationsdesign. Neben der Schätzung von Veränderungen können auch einzelne Zeitpunkte mit dem multivariate GREG Schätzer geschätzt werden. Im Gegensatz zu den bisher etablierten Methoden zur Schätzung von Veränderungen erfordert der multivariate GREG Schätzer keine Imputation der fehlenden Werte aus der Vorperiode für neu hereinrotierende Beobachtungen. Basierend auf einer Monte Carlo Simulation zeigen wir die Effizienzgewinne durch die Verwendung des multivariaten GREG Schätzers im Vergleich zu bisher verwendeten Methoden unter verschiedenen Szenarien.

132 Pendeln und Wohnkosten sparen: Preisgefälle im Stadt-/Umlankontinuum und deren Bewertung im Wohnatlas 2019

Dörte Nitt-Drießelmann

HWWI Hamburgisches Weltwirtschaftsinstitut, Deutschland
nitt-driesselmann@hwwi.org

SECTION: VDS_t

Die räumlich lenkende Wirkung von Kosten/Nutzenüberlegungen zu den Wohn- und der Mobilitätskosten bei Wohnstandortentscheidungen in Städten und ihrem Umland am Beispiel der Big 7 und ihrer Umlandkreise ist Gegenstand einer Deutschland-weiten Analyse des HWWI über die Dörte Nitt-Drießelmann vom HWWI berichtet („Pendeln und Wohnkosten sparen...“).

133 Das Ende der Reurbanisierung? Aktuelle Trends auf dem Wohnungsmarkt der Region Stuttgart

Tobias Held; Attina Mäding

Statistisches Amt der Landeshauptstadt Stuttgart, Deutschland
Tobias.Held@stuttgart.de

SECTION: VDS_t

Mit der provokanten Ansage im Titel „Das Ende der Reurbanisierung...“ nehmen Tobias Heldt und Attina Mäding eine analytische Bewertung aktueller Wohnungsmarkttrends der Region Stuttgart vor.

134 Entwicklungsdynamiken in Großstadtregionen. Eine Betrachtung ausgewählter Indikatoren auf Basis der Innerstädtischen Raumbewertung und der Laufenden Raumbewertung

Jürgen Gödecke-Stellmann; Theresa Lauerbach

Bundesinstitut für Bau-, Stadt- und Raumforschung (BBSR), Deutschland
juergen.goeddecke@bbr.bund.de

SECTION: VDS_t

Der Beitrag von Jürgen Gödecke Stellmann und Teresa Lauerbach vom BBSR nimmt die Entwicklungsdynamiken in Großstadtregionen in den Blick und nutzt dabei eine Datenbasis, die innerstädtische Daten der Kernstädte mit Gemeindedaten der Umlandgemeinden kombiniert. Damit ist es möglich, nicht nur nach Stadt und Umland zu differenzieren, sondern auch die Kernstadt selbst in Lagetypen zu differenzieren.

Näher betrachtet werden sollen ausgewählte Indikatoren und deren Entwicklungsverläufe im Kontinuum von Innenstadt über Stadtrand bis hin zu den näheren und weiteren Verflechtungsbereichen.

135 Soziokulturelle Unterschiede in Großstädten und deren Umland – Ein Entwicklungsvergleich auf der Basis des SOEP

Jan Goebel

DIW Berlin, Deutschland
jgoebel@diw.de

SECTION: VDS

Jan Goebel vom DIW zeigt mittels des SOEP auf, wie sich Bewertungs- und Einstellungsmuster in Großstädten und in Umlandbereichen seit der Jahrtausendwende entwickelt haben. Es wird dargestellt, ob und wenn, in welcher Form sich Verhaltensmuster und Einstellungen bei „Urbanisten“ und „Suburbanisten“ auseinanderentwickelt bzw. angenähert haben.

136 Intensive Profiling – Überblick über einen neuen methodischen Ansatz in der amtlichen Unternehmensstatistik

Simon Rommelspacher (1); Christian Salwiczek (2)

1: Statistisches Bundesamt, Deutschland; 2: Statistisches Amt für Hamburg und Schleswig-Holstein
simon.rommelspacher@destatis.de

SECTION: Business and Market Statistics

Zur Modernisierung der amtlichen Unternehmensstrukturstatistik in Deutschland werden ab dem Bezugsjahr 2018 die Ergebnisse für Unternehmen nach der Definition der EU-Einheitenverordnung (Verordnung (EWG) Nr. 696/93) dargestellt. Nach dieser Definition kann ein Unternehmen aus mehreren Rechtlichen Einheiten bestehen, die gemeinsam eine organisatorische und wirtschaftliche Einheit bilden. Für die Identifizierung und Abgrenzung der Unternehmen in der EU-Definition wird im statistischen Unternehmensregister eine neue Methodik eingeführt, das (Unternehmens-) Profiling. Zur Abgrenzung der Statistischen Unternehmen in großen und komplexen Unternehmensgruppen bzw. Konzernen in Deutschland wird das Intensive Profiling als Methode in den statistischen Verbund eingeführt. Dieses basiert auf einem direkten Austausch mit den zu untersuchenden Konzernen.

Das Statistische Bundesamt hat gemeinsam mit sechs Statistischen Ämtern der Länder von August 2018 bis April 2019 die Methodik und Vorgehensweise entwickelt und in Konzernbesuchen mit großen Unternehmensgruppen in Deutschland getestet. Diese Testphase wurde evaluiert und das Personal in den Statistischen Ämtern für die künftige Durchführung dieser Profiling-Methode geschult. Im Jahr 2019 wird die Methodik „Profiling von Unternehmen“ in den Dauerbetrieb überführt und die Konzernbesuche sollen zu einer guten Qualität der Ergebnisse und zu einer hohen Transparenz der Vorgehensweise beitragen. Die Ergebnisse aus dem Profiling werden für das Berichtsjahr 2018 erstmalig in der amtlichen Unternehmensstrukturstatistik verwendet.

Der Vortrag stellt die Methode Intensive Profiling vor, erläutert die organisatorische Umsetzung der Testphase und des Dauerbetriebs in Deutschland und berichtet im Hauptteil über die Erkenntnisse aus den über zwanzig erfolgten Konzernbesuchen in der Testphase. Hierbei wird u.a. auf die Teilnahmebereitschaft und generelle Resonanz bei den im Fokus stehenden Unternehmen/Konzernen näher eingegangen. Weiterhin wird der Aufwand der Methode vorgestellt und dieser in Relation zum Erkenntnis-Mehrwert des Intensive Profiling gesetzt. Abschließend wird kurz auf die Rechtslage dieser neuen Befragungsform eingegangen und ein kurzer Ausblick für die nächsten Jahre gegeben.

137 Anwendung einer robusten Regression zur Schließung von Datenlücken am Beispiel der Strukturhebung Energie

Maren Koehlmann

Statistisches Bundesamt, Germany
maren.koehlmann@destatis.de

SECTION: Business and Market Statistics

Wichtigste Quelle zur Darstellung der Unternehmenslandschaft in Deutschland sind die Strukturstatistiken, die jährliche Daten zu Investitionen und zur Kostenstruktur der Unternehmen verschiedener Wirtschaftszweige darstellen. Sie sind die zentrale Datenbasis zur Beurteilung der strukturellen Entwicklung von Unternehmen und eine elementare Grundlage für detaillierte Analysen der Wirtschaftspolitik. Zudem können aus den amtlichen Strukturstatistiken wichtige Elemente für die Volkswirtschaftlichen Gesamtrechnungen sowie für umweltökonomische Statistiken gezogen werden. Im Wirtschaftsabschnitt D „Energieversorgung“ und E „Wasserversorgung, Abwasser- und Abfallentsorgung, Beseitigung von Umweltverschmutzungen“ der NACE-Klassifikation der Wirtschaftszweige basieren die Daten auf der Erhebung aller Unternehmen oberhalb bestimmter Abschneidegrenzen. Strukturstatistiken im Produzierenden Gewerbe werden auf der Rechtsgrundlage des „Gesetzes über die Statistik im Produzierenden Gewerbe“ in Verbindung mit dem Bundesstatistikgesetz erhoben.

Eine zweite Datenquelle zur Darstellung wichtiger Merkmale der Unternehmenslandschaft basiert auf Auswertungen von Verwaltungsdaten. Sie speisen sich auf rechtlicher Grundlage des Verwaltungsdatenverwendungsgesetzes aus dem Unternehmensregister, in welchem wiederum Umsatzangaben von den Finanzverwaltungen sowie Beschäftigtendaten von der Bundesagentur für Arbeit zusammenfließen.

Aufgrund der gesetzlichen Vorgabe kleine Unternehmen bürokratisch zu entlasten, gibt es bei der Strukturstatistik Energie festgelegte Abschneidegrenzen. Beispielsweise haben im Bereich der Energieversorgung nur Unternehmen eine Auskunftspflicht, deren Jahresumsatz 1 Million € (WZ 35.3) bzw. 3 Millionen € (WZ 35.1 und 35.2) und mehr beträgt. Dies führt in der Konsequenz dazu, dass die Ergebnisse auf Basis der Erhebung zur Unternehmensstruktur insbesondere beim Merkmal Anzahl der Unternehmen erheblich von den Ergebnissen des Unternehmensregisters abweichen. Um diese Inkonsistenz zu beheben, und die Datenlücken bei den Kleinunternehmen zu schließen, wird eine Datenergänzung für die Merkmale der Strukturhebung Energie vorgenommen. Dabei ist zwischen Merkmalen zu unterscheiden, die direkt aus Verwaltungsdaten übertragen werden können (z.B. Umsatz) und denjenigen, die im Erhebungsprogramm der Strukturstatistik enthalten sind, aber nicht in den Verwaltungsdaten. Für diese müssen entsprechende Schätzmodelle entwickelt werden.

Im Vortrag werden Vorgehensweise und Ergebnisse der Schätzungen präsentiert. Dabei werden unterschiedliche Schätzmethoden miteinander verglichen.

138 A new non-monotone link function for marketing analysis

Gloria Gheno

Innovative data analysis, Italy
gloriagheno@libero.it

SECTION: Business and Market Statistics

In the last decades the most important companies of the technological sector announce the new products in a spectacular way to intrigue and induce the media to try to predict the news coming out. In this work I analyze how current and potential customers, based on their specific characteristics, are interested in the leaked information, in particular considering, for example, the technologic products owned by them and their brand loyalty. Indeed, for a company knowing the target customers interested in launch of new technologies is essential to be able to develop products which reflect their needs so as to fully meet their expectations. To study the degree of interest in the advertising campaign, a month before the launch I ask a sample of people with which probability it assumes that a new product with particular characteristics is launched. Then, I compare these results with the product actually put on the market. I divide the study into two parts, in the first part I analyze how the subjects form their own opinions, in the second part I evaluate the goodness of the forecast. I study the factors which affect these probabilities using the Beta regression and a link function which I created specifically to eliminate the typical monotonicity of the traditional link functions. I demonstrate the real goodness of this new technique by comparing it with others already present in the literature using specific statistical tests. This work offers companies in the technology sector the opportunity to save time and money by eliminating the costs of developing unrequired characteristics.

139 Welche Arbeitsagentur ist „auffällig“? Zum Vergleich von Performance-Indikatoren

Malte Schierholz; Tobias Büttner

Bundesagentur für Arbeit, Germany
malte.schierholz@iab.de

SECTION: Business and Market Statistics

Zum Erreichen ihrer geschäftspolitischen Ziele werden in der Bundesagentur für Arbeit Zielvereinbarungen mit jeder Arbeitsagentur abgeschlossen. Ein zentrales Ziel ist beispielsweise, möglichst viele Arbeitslose in Arbeit zu vermitteln, was anhand des Indikators „Integrationsquote“ gemessen wird. Die Zielerreichung bei den verschiedenen Indikatoren wird monatlich geprüft. Bei auffälligem Abschneiden finden so genannte anlassbezogene Management-Dialoge zur Klärung des Sachverhalts statt.

Die statistische Herausforderung besteht darin, anhand einer Vielzahl vorhandener Indikatoren einzelne Arbeitsagenturen zu identifizieren, die in diesen Dialogen thematisiert werden sollen. Ein direkter Vergleich der Arbeitsagenturen anhand ihrer Indikatoren wäre wenig aussagekräftig, da für Einflussfaktoren wie die lokale Arbeitsnachfrage oder Strukturmerkmale der betreuten Arbeitslosen kontrolliert werden muss.

Im Vortrag werden Lösungsansätze diskutiert und ein Bezug zu Methoden der kausalen Inferenz wird hergestellt.

140 Model averaging in non-nested models with an application to optimal designs

Kira Alhorn (1); Holger Dette (2); Kirsten Schorning (2)

1: TU Dortmund, Germany; 2: Ruhr-Universität Bochum, Germany
kira.alhorn@tu-dortmund.de

SECTION: Statistical Theory and Methods

We consider model averaging estimators in the case where the competing models are non-nested and none of these models is correctly specified. We derive the asymptotic distribution of the model averaging estimate with fixed weights and study its finite sample properties via simulations. Many results presented in literature indicate that model averaging has some advantages over model selection. We demonstrate that conclusions of this type depend sensitively on the class of models under consideration. In particular we observe some advantages of estimation after model selection if the competing models are of rather different shape. We use the asymptotic results to construct optimal designs minimizing an expectation of the asymptotic mean squared error of the model averaging estimate. We demonstrate that these designs can improve the accuracy of model averaging estimators substantially. Moreover, the derived designs also improve the accuracy of estimators in a model selected by model selection and model averaging estimates with random weights.

141 Survey-weighting of Linear Mixed Models under Box-Cox and Dual Transformations

Jan Pablo Burgard; Patricia Dörr

Universität Trier, Deutschland
doerr@uni-trier.de

SECTION: Statistical Theory and Methods

In econometrics, the analysis of positive and skewed variables such as income, wealth or returns of sales is of central concern. These data often stem from a complex survey design, which can lead to biased regression estimators if the design not accounted for. Furthermore, data often inherit a clustered structure that finds expression in a multilevel model, for example for panels and small area estimation.

When skewness persists even after conditioning on explanatories in this context, regression analysis encounters a threefold challenge: First, a data transformation is required in order to reestablish the validity of model assumptions. Common transformations are dual and Box-Cox transformations, that dispose of a free parameter and the choice thereof should be data driven. Second, the cluster structure that pertains the multilevel set-up should be respected. Third, accountance of the survey design is essential for consistent estimation of all parameters of the statistical model.

We propose a flexible algorithm that incorporates survey weights in a mixed effects regression model under transformations. The transformation can be chosen from the family of Box-Cox or dual transformations and the transformation parameter is estimated using the available data.

142 Large sample properties of an IV estimator based on the Ahn and Schmidt moment conditions

Andrew Adrian Yu Pua (1); Markus Fritsch (2); Joachim Schnurbus (2)

1: Xiamen University, China; 2: University of Passau, Germany
joachim.schnurbus@uni-passau.de

SECTION: Statistical Theory and Methods

We propose an instrumental variables (IV) estimator based on nonlinear (in parameters) moment conditions for estimating linear dynamic panel data models and derive the large sample properties of the estimator. We assume that the only explanatory variable in the model is one lag of the dependent variable and consider the setting where the absolute value of the true lag parameter is smaller or equal to one, the cross section dimension is large, and the time series dimension is either fixed or large. Estimation of the lag parameter involves solving a quadratic equation and we find that the lag parameter is point identified in the unit root case; otherwise, two distinct roots (solutions) result. We propose a selection rule that identifies the consistent root asymptotically in the latter case and derive the asymptotic distribution of the estimator for the unit root case and for the case when the absolute value of the lag parameter is smaller than one.

143 Practical aspects of using nonlinear moment conditions in linear dynamic panel data models

Andrew Adrian Yu Pua (1); Markus Fritsch (2); Joachim Schnurbus (2)

1: Xiamen University, China; 2: University of Passau, Germany
markus.fritsch@uni-passau.de

SECTION: Statistical Theory and Methods

We study the estimation of the lag parameter of linear dynamic panel data models with first order dynamics based on nonlinear (quadratic) moment conditions. Our contribution is twofold: First, we show that extending the standard assumptions by mean stationarity and time series homoscedasticity and employing these assumptions in estimation restores standard asymptotics and mitigates the non-standard distributions found in the literature. Second, we consider an IV estimator based on the quadratic moment conditions that consistently identifies the true population parameter under standard assumptions. Standard asymptotics hold for the estimator when the cross section dimension is large and the time series dimension is finite. We also suggest a data-driven approach to obtain standard errors and confidence intervals that preserves the time series dependence structure in the data.

144 Maschinelles Lernen in den Handwerksstatistiken

Joerg Feuerhake

Statistisches Bundesamt, Germany
joerg.feuerhake@destatis.de

SECTION: Methodology of Statistical Surveys

Im Jahr 2017 wurde von Beck und Dumpert ein Verfahren aus den Handwerksstatistiken vorgestellt, bei dem eine Support Vector Machine (SVM) verwendet wird, um irrelevante Fälle in der Grundgesamtheit der Handwerksunternehmen zu identifizieren. So sollten durch den Einsatz maschineller Lernverfahren personelle Ressourcen entlastet und gezielter eingesetzt werden.

Bei dem Verfahren werden Daten aus Vorperioden genutzt, um eine SVM zu trainieren. Um, bei der vorliegenden Anzahl an Variablen und Beobachtungen, ein Tuning und Training der SVM mit den vorhandenen Ressourcen in akzeptabler Zeit zu realisieren, muss die Zahl der Merkmale auf die relevantesten reduziert werden. Für diese Feature Selection wird ein Random-Forest-Algorithmus eingesetzt.

Nach nun mehrjährigem Einsatz des Verfahrens kann über den Stand bei der Überführung in den produktiven Betrieb, den produktiven Betrieb selbst und die dabei gemachten Erfahrungen berichtet werden. Zusätzlich werden mögliche Weiterentwicklungen umrissen.

145 Maschinelles Lernen zur Imputation: Eine kritische Betrachtung

Florian Dumpert

Statistisches Bundesamt, Germany
florian.dumpert@destatis.de

SECTION: Methodology of Statistical Surveys

Neuen statistischen Methoden sollte immer die Chance gegeben werden, sich auch in der amtlichen Statistik zu beweisen – aber nicht zum Selbstzweck, sondern um einen Mehrwert zu generieren. Verfahren, die in der einschlägigen Literatur zum sogenannten maschinellen Lernen gerechnet werden, wurden bislang hauptsächlich zum Zweck der Klassifikation in der amtlichen Statistik erprobt. Über verschiedene Anwendungsoptionen wurde im Rahmen vergangener Statistischer Wochen bereits berichtet.

Eine andere statistische Fragestellung betrifft den Bereich der Imputation, also das Ersetzen von fehlenden und unplausiblen Werten in einem vorliegenden Datenmaterial mit dem Ziel, die wahre, den Daten zugrunde liegende Verteilung abzubilden. Motiviert durch den häufig gewinnbringenden Einsatz von Methoden des maschinellen Lernens bei Fragen der Klassifikation und Regression sollten solche auch als Komponente von Imputationsschritten erprobt werden. Hierzu wurde im Statistischen Bundesamt eine Studie mit amtlichen Daten durchgeführt. Der Vortrag berichtet über die Ergebnisse der Studie und bewertet diese unter für die amtliche Statistik relevanten Gesichtspunkten.

146 Einführung einer automatisierten Plausibilitätskontrolle in der neuen digitalen Verdiensterhebung (NVE)

Beck, Martin; Mätzig, Paul

Statistisches Bundesamt, Germany
martin.beck@destatis.de

SECTION: Methodology of Statistical Surveys

Im September 2018 hat das Bundesministerium für Wirtschaft und Energie das Statistische Bundesamt mit der Erstellung eines Konzepts zur Verzahnung der vierteljährlichen Verdiensterhebung und der vierjährigen Verdienststrukturerhebung zu einer neuen digitalen Verdiensterhebung (NVE) beauftragt. Damit sollen neu aufgetretene Informationsbedarfe in wichtigen politischen Handlungsfeldern gedeckt werden, so z.B. mit Blick auf die Evaluation des Mindestlohnes oder die Analyse geschlechtsspezifischer Verdienstunterschiede (Gender Pay Gap).

Das von den statistischen Ämtern entwickelte Konzept sieht vor, die NVE ab Januar 2021 monatlich bei 60 000 Betrieben durchzuführen und dabei ausschließlich auf Daten zurückzugreifen, die für andere gesetzlich angeordnete Zwecke, z.B. die Erstellung von Lohnbescheinigungen, ohnehin in den Betrieben vorliegen bzw. ohne größeren Aufwand zur Verfügung gestellt werden können.

Im Rahmen der NVE werden von den Statistischen Ämtern der Länder monatlich ca. 7 Mio. Arbeitnehmerdatensätze verarbeitet werden müssen. Die erforderliche Überprüfung dieser Daten auf fehlerhafte oder fehlende Angaben und deren Korrektur, d.h. die so genannte Plausibilitätskontrolle, kann angesichts der Datenmenge und der engen Terminvorgaben einer monatlichen Statistik nicht wie bisher manuell erfolgen. Es besteht die Notwendigkeit einer (weitgehend) automatisierten Plausibilitätskontrolle. Im Statistischen Bundesamt wurde daher in einem Proof of Concept untersucht, ob und inwieweit eine solche automatisierte Plausibilitätskontrolle in der neuen digitalen Verdiensterhebung realisierbar ist und inwiefern hierbei Machine-Learning-Verfahren hilfreich sind. Im Fokus standen zunächst die Verfahren CANCEIS und HoloClean, deren Eignung anhand der Daten der Verdienststrukturerhebung 2014 getestet wurde. Während CANCEIS sich bereits als geeignet erwiesen hat, sollen die Tests von HoloClean und ggf. weiteren Machine-Learning-Verfahren, wie z.B. MacroPCA, noch fortgeführt werden.

Im vorliegenden Beitrag präsentieren wir die Vorgehensweise und die Ergebnisse des Proof of Concept und der weitergehenden Untersuchungen zur Einsetzbarkeit einer automatisierten Plausibilitätskontrolle in der geplanten NVE.

147 „Es liegt nicht immer nur am Empfänger“ – Kommunikation statistischer Ergebnisse durch Statistiker

Tanja Ihden

Universität Bremen, Germany
tanja.ihden@uni-bremen.de

SECTION: Statistical Literacy

In der Debatte um eine „Statistical Literacy“ der Bevölkerung im privaten wie professionellen Kontext wird häufig betont, dass der „Konsument“ bzw. „Empfänger“ statistischer Information über eine gewisse statistische Grundbildung verfügen müsse, um Auswertungen von Zahlen adäquat interpretieren zu können. Eine „Statistical Literacy“ sollte allerdings nicht nur auf der Seite der Empfänger statistischer Information gefördert werden. Häufig erscheint die Kommunikation statistischer Ergebnisse durch Statistiker sperrig, unzugänglich, zuweilen sogar unverständlich. Dabei sollte der Statistiker als „Sender“ von Informationen es nicht nur als seine Aufgabe verstehen, Daten auszuwerten, sondern auch, diese angemessen aufbereitet an den „Empfänger“ weiterzugeben. Dabei sollte bedacht werden, dass diese mit unterschiedlichem Vorwissen und verschiedenen Assoziationen ausgestattet sind. Der Vortrag arbeitet heraus, wie Statistiker ihre Ergebnisse kommunizieren können, damit diese auf der Seite der Empfänger das Verständnis erleichtern. So sollte zum einen bedacht werden, dass ein unterschiedliches Verständnis von scheinbar allgemein gleich definierten Begriffen vorliegen kann (z.B. Begriff der „Signifikanz“). Zum anderen wird die Überlegung angestellt, ob die Präsentation statistischer Ergebnisse im Rahmen eines „Storytelling“ nicht erheblich zu einer Steigerung des Verständnisses führt. Der Statistiker sollte zudem stets überdenken, auf welche Weise die Daten einfach, übersichtlich und nachvollziehbar dargestellt werden. Dazu gehört auch zu hinterfragen, was der Empfänger wissen MUSS, um die Kernaussagen der statistischen Auswertung nachzuvollziehen. Zudem wird im Vortrag betont, dass nicht nur die Ergebnisse, sondern auch die Methodik der Erhebung bzw. Auswertung statistischer Daten visualisiert werden könnte. Da Bilder als universelle Sprache gelten, entfällt hier auf der einen Seite die Problematik des unterschiedlichen Vokabulars, während auf der anderen Seite Bilder häufig nicht nur schneller, sondern auch leichter erfasst werden als die textliche Form. Letztlich soll im Rahmen des Vortrages ein Verständnis dafür geschaffen werden, dass statistische Information abstrakt ist und somit bei der Vermittlung dieser bewusst vorgegangen werden sollte.

148 Data Literacy: Schlüsselkompetenz des 21. Jahrhunderts

Katharina Schüller

STAT-UP Statistical Consulting & Data Science GmbH, Deutschland
katharina.schueller@stat-up.com

SECTION: Statistical Literacy

Welches Wissen, welche Fähigkeiten, welche Haltung benötigt es in Gesellschaft, Arbeitswelt und Wissenschaft, in denen Daten als wertvolle, mitunter die wertvollste Ressource gelten und Entscheidungen zunehmend auf der Grundlage von Daten getroffen werden? Zweifellos werden Digitalisierung und

Datafizierung das Leben und Arbeiten im 21. Jahrhundert nachhaltig verändern. Künstliche Intelligenz, vernetzte Produktion, kommunizierende Maschinen und selbstfahrende Autos werden von Daten gesteuert und produzieren selbst Daten am laufenden Band. Daten sind die Ausgangsbasis für Wissens- bzw. Wertschöpfung als Grundlage für bessere Entscheidungen.

Um systematisch Wissen bzw. Wert aus Daten zu schöpfen, ist deshalb zukünftig in allen Sektoren und Disziplinen die Fähigkeit, planvoll mit Daten umzugehen und sie im jeweiligen Kontext bewusst einsetzen und hinterfragen zu können, von entscheidender Bedeutung. Dies wird als Data Literacy bezeichnet und umfasst die Fähigkeiten, Daten auf kritische Art und Weise zu sammeln, zu managen, zu bewerten und anzuwenden. Data Literacy ist weit mehr ein breites und tiefes Detailwissen über sich laufend verändernde Methoden und Technologien. Vielmehr spielt die Dimension der Datenethik, der Motivation und Werthaltung eine zentrale Rolle, um zukünftig mit Daten erfolgreich und souverän umgehen zu können.

Worum es sich bei Data Literacy als „Schlüsselkompetenz des 21. Jahrhunderts“ genau handelt, ist Gegenstand aktueller Forschung. Im Auftrag des Hochschulforums Digitalisierung (HFD), einer gemeinsamen Initiative des Stifterverbandes für die Deutsche Wissenschaft mit dem CHE Centrum für Hochschulentwicklung und der Hochschulrektorenkonferenz, haben wir den Begriff geschärft und erstmals einen umfassenden Data Literacy Kompetenzrahmen entwickelt. Er eignet sich zur Ableitung von Lernzielen in zahlreichen Fachgebieten und Studiengängen. Mögliche Mess- und Testverfahren für Data Literacy werden näher beleuchtet. Beispielhaft ist anhand von Fallstudien aufgezeigt, wie der Kompetenzrahmen in realen Problemsituationen dazu beitragen kann, Daten nutzbar zu machen. Dabei spielt insbesondere die Dimension der Datenethik, der Motivation und Werthaltung eine zentrale Rolle.

149 Separierung von demografischen und verhaltensbedingten Einflüssen auf die Entwicklung der Lebensformen

Tim Hochgürtel

Statistisches Bundesamt, Germany
tim.hochguertel@destatis.de

SECTION: Statistical Literacy

Im Verlauf der vergangenen zwei Dekaden hat sich das Verhältnis von familiären Lebensformen (Paare mit Kindern, Alleinerziehende) zu nicht-familiären Lebensformen (Paare ohne Kinder, Alleinstehende) verschoben. Während sich die Anzahl der Familien insgesamt reduziert hat, ist die Anzahl der Paare ohne Kinder, besonders aber die Anzahl der Alleinstehenden gestiegen.

Aus der Entwicklung der absoluten Zahlen ist nicht ersichtlich, ob diese Veränderung vorrangig demografisch oder verhaltensbedingt ist. Die Anzahl der Menschen, welche sich in einem Alter befinden, in dem typischerweise Familien gegründet werden, ist rückläufig. Daher ist aus demografischen Gegebenheiten heraus ein Rückgang der Familien zu erwarten. Folgt man hingegen den Thesen einer Individualisierung der Gesellschaft, erscheinen auch verhaltensbedingte Veränderungen als Ursache plausibel. Die Gründung einer Familie wird für Menschen mehr und mehr von normativen Erwartungen entlastet und zu einer biografischen Option neben anderen sozial akzeptierten Alternativen.

Um zu quantifizieren, in welchem Maße demografische und verhaltensbedingte Faktoren die Veränderungen im Verhältnis von familiären und nicht-familiären Lebensformen verursachen, wird auf eine Standardisierung der demografischen Struktur zurückgegriffen. Durch eine Anpassung der Gewichtung

des Mikrozensus wird die Alters- und Geschlechtsverteilung eines Referenzjahres in ein Vergleichsjahr projiziert. Somit lässt sich die Veränderung der Lebensformen im Vergleichsjahr gegenüber dem Referenzjahr unter Kontrolle demografischer Einflüsse betrachten. Ein Vergleich der empirischen Veränderung und der ermittelten Veränderung unter Kontrolle demografischer Einflüsse erlaubt die Trennung von demografischen und verhaltensbedingten Einflussfaktoren.

150 Asset Pricing using Time-Frequency Dependent Network Centrality

Jozef Barunik (1); Michael Ellington (2)

1: Charles University in Prague, Institute of Economic Studies, Tschechische Republik; 2: University of Liverpool
barunik@fsv.cuni.cz

SECTION: Statistics in Finance

We provide a framework where time-frequency dependent network centrality links to expected excess return of financial assets. Noting that investors trade on different horizons, it is essential to understand how the connectedness of a system influences risk premium over the short-, medium- and long-term. Viewing the market as being generated by a time-varying parameter VAR model implies that a shock to the j -th asset is time-frequency dependent. This creates a network of time-frequency connections among all assets in the market. We propose a new measure of time-frequency dependent network centrality and apply this to all stocks listed on the S&P500. Our findings indicate that our time-frequency dependent measures significantly price assets; particularly over the longer-term.

151 Intraday Conditional Value at Risk: A Periodic Mixed-Frequency GAS Approach

Bastian Gribisch; Tobias Eckernkemper

Universität zu Köln, Deutschland
bastian.gribisch@statistik.uni-koeln.de

SECTION: Statistics in Finance

We propose a copula-based periodic mixed frequency GAS framework in order to model and forecast the intraday Exposure Conditional Value at Risk (ECoVaR) for an intraday asset return and the corresponding market return. In particular we analyze GAS models which account for long-memory-type of dependencies, periodicities, asymmetric nonlinear dependence structures, fat-tailed conditional return distributions and intraday jump processes for asset returns. We apply our framework in order to analyze the in-sample and out-of-sample ECoVaR forecasting performance for a large data set of intraday asset returns of the S&P500 index.

152 Non-Parametric Prediction of Financial Risk using Artificial Neural Networks. Does High Frequency Information Help?

Christian Muecher

Universität Konstanz, Germany
christian.muecher@uni-konstanz.de

SECTION: Statistics in Finance

Existing papers on non-parametric modelling of conditional variance functions using Neural Networks rely on daily returns as inputs. This project aims to use high frequency returns of financial assets as inputs for Neural Networks in order to learn the conditional variance function.

Depending on the frequency the intra-day data is sampled at the resulting input vector becomes very high dimensional. E.g. for data sampled at the 5 minute frequency this results to 390 intra-day returns for each day. Bottleneck Neural Networks are able to deal with high-dimensional input data. The question is whether the additional information contained in the high frequency data improves the performance of NN based non-parametric estimates of the conditional variance function.

The availability of high-frequency financial data, sampled intra-daily at different frequencies (5 minutes, 1 second, 1 nanosecond, ...) gave rise to models that use »Realized Volatility« (RV), an estimator for the daily volatility obtained from summing up the squared intra-day returns of that day and taking the square root. A particularly popular model is the Heterogeneous Autoregressive Model (HAR) which aggregates intra-day information to a daily measure that is used as an input to predict daily volatility.

The HAR is a parametric linear model. Using Neural Networks I am able to use unprocessed, raw financial returns data to non-parametrically estimate non-linear relationships to predict financial risk. I show that, using the same high-frequency information as in the HAR, the predictive ability of the Neural Network based model outperforms that of the HAR model.

153 A proposal to assess the representativeness of non-probability surveys

Alessandra Petrucci; Emilia Rocco

Università degli Studi di Firenze, Italien
alessandra.petrucci@unifi.it

SECTION: Joint DStatG & SIS – Survey Statistics

Probability sampling remains the gold standard for survey research. However, during the past years we have registered an increasing decline of the response rates in the „traditional“ probability surveys and, at the same time, a proliferation of non-probability source of data. With recent technological innovations, it is increasingly convenient and cost-effective to collect large numbers of highly non-presentative samples via online surveys. At the same time, due to the low response rate the quality advantage of the more expensive probability surveys is increasingly declining. The main problem caused by non-representative survey data is that estimators of population characteristics must be assumed to be biased unless convincing evidence to the contrary is provided.

Therefore quality indicators need in order to assess their quality and their potential for preserving a valid statistical inference framework. The aim of the present work is to investigate the possibility to use for assessing the quality of non-probability source of data, some quality indicators suggested in literature for evaluating the risk of non-response bias. In both cases (non-response and non-probability sampling) it must be assumed that the potential bias in the data is the result of a self-selection process, but in the two cases the data collection process may be very different and the information on the data collection process are usually of great relevance in order to evaluate their quality.

Moreover, whenever the data are observed only on a non-probability subset of the population the potential bias of the statistic of interest cannot be assessed except through indirect measures based on more or less reasonable assumptions and on the use of data external to the survey. What these assumptions may be and which the auxiliary variables available for the different non-probabilistic sources is a relevant question in the current context.

154 Techniques for randomising protest survey samples to facilitate comparison with nationally representative samples

Natalie Shlomo (1); Clare Saunders (2)

1: University of Manchester; Vereinigtes Königreich; 2: University of Exeter, Vereinigtes Königreich
Natalie.shlomo@manchester.ac.uk

SECTION: Joint DStatG & SIS – Survey Statistics

In this paper, we compare two techniques for balancing non-random protest survey samples with random European Social Survey (ESS) samples of protesters. These techniques will become increasingly important in the study of protesters because protesters are a „rare“ sample. The power for analysis

is very low if we rely on random samples of protesters alone. Therefore, it is important to find ways in which to randomise data from other data sources to provide rich insights into protesters. The two techniques compared are proportional weighting and sample matching. Proportional weighting involves post-stratification bench-marking based on UK census data. Sample matching is an approach modified from quasi-experimental designs. Propensity score matching is used to balance non-random samples. We show that that the sample matching approach to induce randomization of the non-random sample of protesters through the integrations with the ESS has the highest standard of replication of the distribution of variables in a random sample of protesters. We are able to correct for under-selection of centrists and over-selection of left-wingers, especially among the stalwart protesters.

155 Bias correction for non-probability samples

Ralf Münnich; Simon Lenau

Universität Trier, Deutschland
muennich@uni-trier.de

SECTION: Joint DStatG & SIS – Survey Statistics

Web surveys are used more and more to gather data in the social sciences. Often, the data generating process is largely unknown. Obviously, classical concepts of inference immediately raise concerns about the biasedness of estimates from such kinds of non-probability samples. The aim of the presentation is to give a short overview of different approaches to compensate for the non-representativeness. Pseudo-design-based approaches focus on weighting data in order to handle non-probability samples in ways analogous to classical probability samples. These methods mainly include propensity score and calibration weighting, as well as extensions and combinations of both. On the other hand, model-based approaches concentrate on predicting the variable(s) of interest from known auxiliary data. A large variety of methods is discussed in this context, ranging from classical regression models to machine learning algorithms. A comparative study using Monte-Carlo simulation methods is conducted to evaluate the appropriateness of the methods in a practical environment. Finally, the methods will be discussed in an application to the WageIndicator Web Survey, which uses a self-selection strategy to recruit respondents and is thus likely.

156 Das Kodiersystem Iris/MUSE – Basis der multi- und unikausalen Todesursachenstatistik in Deutschland

Olaf Eckert

Statistisches Bundesamt, Germany
olaf.eckert@destatis.de

SECTION: DGD

In Deutschland nutzen sieben von 14 Statistischen Landesämtern das elektronische Kodiersystem der Todesursachenstatistik Iris/MUSE vollumfänglich. Voraussetzung für die Nutzung des Systems ist die vollständige elektronische Erfassung von Todesbescheinigungen, die in der Regel in Gesundheitsämtern erfolgt. Die fallweise Bearbeitung geschieht in drei Schritten:

- 1) ICD-10-Kodierung medizinischer Diagnosen: Nach einem automatischen Kodierversuch der integrierten Texterkennung unter Nutzung eines vom DIMDI bereitgestellten Wörterbuches werden unkodierte Textteile von den Signierfachkräften nachkodiert.
- 2) Ermittlung der multikausalen Todesursachen
- 3) Selektion des Grundleidens der Todesursachenstatistik

Die Schritte 2 und 3 führt der Kodierkern „Multicausal and Unicausal Selection Engine“ (MUSE) durch, der vom Statistischen Bundesamt entwickelt wurde und in der internationalen Iris-Version seit Anfang 2017 den langjährigen US-amerikanischen Standard ACME/MICAR ersetzt. Die Kodierentscheidungen von MUSE beruhen auf einem medizinisch-fachlichen Regelwerk der Mortalitätsstatistik mit fast 170.000 Detailanweisungen. Diese multi- und unikausalen Entscheidungstabellen werden von der Iris Core Group jährlich nach den Vorgaben der ICD-10 aktualisiert.

Es werden multi- und unikausale Ergebnisse und Kennzahlen für die Datenqualität für anonymisierte, nicht-repräsentative Daten (n=5.800) berechnet und präsentiert.

Aus methodischer Sicht ist ein flächendeckender Umstieg der Statistischen Landesämter auf das Kodiersystem Iris/MUSE geboten. Das Regelwerk der ICD-10 zur Selektion des Grundleidens und zur multikausalen Kodierung ist komplex und kann nur mit elektronischer Unterstützung fachgerecht und effizient angewandt werden. Systematischen Verzerrungen der Todesursachenstatistik könnte durch eine einheitliche Anwendung des Regelwerkes entgegengewirkt werden. Die Anforderungen an eine multikausale Kodierung der Todesursachen werden zudem mit der Einführung der ICD-11 erheblich steigen.

Trotz der Fortschritte bei der elektronischen Signierung gibt es weiterhin gravierende Qualitätsprobleme ärztlicher Todesbescheinigungen, wie fehlende Informationen oder fehlerhafte Kausalketten. Nach den Kriterien der WHO waren 10,7% der Todesursachen der deutschen Todesursachenstatistik 2015 schlecht definiert.

Eine künftige bundeseinheitliche elektronische Todesbescheinigung (eTB), die bei Dokumentationsproblemen Warnhinweise für Ärztinnen und Ärzte generiert und die Daten an Gesundheitsämter übermittelt, könnte zu einer Verbesserung der Datenqualität führen.

157 Ungültige Codes in der Todesursachenstatistik, Definition und regionale Unterschiede

Annelene Wengler (1); Heike Gruhl (2); Dietrich Plaß (2); Janko Leddin (1); Elena von der Lippe (1)

1: Robert Koch-Institut, Deutschland; 2: Umweltbundesamt
wenglera@rki.de

SECTION: DGD

Die Anzahl der verlorenen Lebensjahre (Years of Life Lost, YLL) ist ein zentraler Indikator bei Krankheitslaststudien nach Vorbild der Global Burden of Disease-Studie (GBD). YLL entsprechen hierbei der ferneren Lebenserwartung einer Person zum Todeszeitpunkt. Die Analyse der YLL erfolgt todesursachenspezifisch. Die Grundlage für die Berechnung der YLL im BURDEN 2020-Projekt, stellt die Todesursachenstatistik des Statistischen Bundesamts (Destatis) dar.

Bei der Erfassung und Verarbeitung der Daten der Todesursachenstatistik kann es zu Fehlern kommen. Es treten sogenannte ungültige ICD-10-Codes auf, die keine ursächliche Todesursache beschreiben. Dies sind Ursachen, die nicht unmittelbar zum Tod führen (z.B. Herzinsuffizienz), Symptome von Erkrankungen (z.B. Rückenschmerzen), unplausible Ursachen (z.B. Hodenkrebs bei Frauen), unspezifische Ursachen (z.B. unspezifische Krebserkrankung) oder Folgeerkrankungen (z.B. Nierenversagen). Einem Teil der Todesfälle kann damit nicht zweifelsfrei eine Todesursache zugeordnet werden.

Bei der Definition von ungültigen Codes folgen wir der Systematik der GBD-Studie (Roth et al. 2018). In der GBD-Studie werden darüber hinaus Methoden zum Umgang mit ungültigen Codes angewandt. Auch im Rahmen des Projekts BURDEN 2020 wird die Qualität der Daten nach Anzahl, Art und geographischer Verteilung ungültiger Codes analysiert. Schlussendlich erfolgt eine Umverteilung der Sterbefälle mit ungültigen auf Fälle mit gültigen Codes.

In Deutschland weisen 2015 27% der Sterbefälle einen ungültigen Code auf, mit starken Varianzen zwischen den Bundesländern. So liegt der Anteil in Sachsen-Anhalt bei 16%, in Bremen jedoch bei 35%. Diese Differenzen können teilweise durch die unterschiedliche Erfassung und Verarbeitung der Daten in den Bundesländern erklärt werden. So wird in einigen Bundesländern bereits das Kodiersystem Iris/MUSE eingesetzt, um den jeweils richtigen Code für das Grundleiden zu bestimmen. Die Liste der als ungültig klassifizierten Codes der GBD-Studie unterscheidet sich dabei von denen, die von WHO und Destatis verwendet werden. Der Beitrag zeigt eine Gegenüberstellung und diskutiert das Für und Wider der nachträglichen Adjustierung der Todesursachenstatistik.

158 Ergebnisse eines Eurostat-Projektes zur Modernisierung der europäischen Produktionsstatistik

Luis-Federico Flores; Luisa Baumgärtner

Statistisches Bundesamt, Deutschland
luisa.baumgaertner@destatis.de

SECTION: Business and Market Statistics

Im Rahmen eines Eurostat-Projektes untersucht das Statistische Bundesamt verschiedene Methoden zur Modernisierung der europäischen Produktionsstatistik (PRODCOM). Die europäische Produktionsstatistik ist eine jährliche Statistik, deren Ergebnisse für Deutschland aus den Daten der monatlichen und vierteljährlichen Produktionserhebung ermittelt werden. Diese Daten beinhalten Menge und Wert der im Inland produzierten Güter aller Unternehmen des Verarbeitenden Gewerbes. Allerdings nur für diejenigen Unternehmen mit 20 oder mehr Beschäftigten.

Um auch die Unternehmen mit weniger als 20 Beschäftigten in die Ergebnisse mit einzubeziehen – ohne jedoch die Unternehmen zusätzlich zu belasten – soll die Produktion der Unternehmen mit weniger als 20 Beschäftigten (Kleinstunternehmen) auf Einzeldatenebene mit Hilfe von Verwaltungsdaten geschätzt werden. Der Beitrag berichtet über die Ergebnisse zur Schätzung des Produktionswertes für Kleinstunternehmen anhand verschiedener Regressionsmodelle, mit Hilfe derer der mathematische Zusammenhang zwischen Produktionswert und bereits vorhandenen Merkmalen aus den Verwaltungsdaten modelliert werden kann. Die Produktionsstatistik nutzt somit die Potentiale von Verwaltungsdaten für eine belastungsarme Erstellung der Ergebnisse bei gleichzeitiger Optimierung der Informationen für die Datennutzer.

Aufgrund der kleinteiligen Darstellung der Ergebnisse nach der europäischen Güterklassifikation (die sog. PRODCOM-Liste) kommt es sehr häufig zu einer Sperrung bzw. Geheimhaltung von Ergebnissen, da nur wenige Unternehmen diese Güter produzieren oder es dominierende Unternehmen gibt. Daher geht es in einem zweiten Teil des Projektes um die mögliche Reduzierung von geheimzuhaltenden Ergebnissen. Zwei Maßnahmen hierzu wurden im Rahmen des Projektes getestet und werden im Beitrag vorgestellt: Eine Maßnahme beinhaltet die Reduzierung des sogenannten p-Wertes von aktuell 5% (=Schwellenwert zur Bestimmung von dominanten Unternehmen) und die damit verbundene Offenlegung von bisher geheimzuhaltenden Ergebnissen der nationalen und europäischen Produktionsstatistik. Eine zweite Maßnahme beinhaltet eine Befragung zum Testen der Bereitschaft der Unternehmen zur Freigabe ihrer ansonsten geheimzuhaltenden Daten. Abgedeckt durch den § 16 Abs. 1, Nr. 1 BStatG kann das Statistische Bundesamt Ergebnisse, die eigentlich der Geheimhaltung unterliegen, veröffentlichen, sofern die betroffenen Unternehmen zustimmen.

159 Zirkuläre Wirtschaft – Herausforderungen für Industrie und Statistik

Angelika Becker

Verband der Chemischen Industrie e.V., Deutschland
becker@vci.de

SECTION: Business and Market Statistics

Die Europäische Kommission hat Ende 2015 ein Paket zur „Circular Economy“ vorgelegt, um Wirtschaft und Konsum nachhaltiger zu gestalten. Seit dem nimmt das Thema Zirkuläre Wirtschaft immer mehr an Fahrt auf.

Für die Chemie ist die zirkuläre Wirtschaft dabei weit mehr als nur Recycling. Das Konzept der zirkulären Wirtschaft umfasst alle Beiträge zur Schonung von Ressourcen. Zentral ist die Steigerung der Effizienz zur Minimierung des Ressourcenbedarfs über den gesamten Produktlebenszyklus hinweg. Bei der Produktion anfallende Nebenprodukte werden wenn möglich wieder direkt in der Produktion eingesetzt. Am Produktlebensende kann die Chemie mehrere Kreisläufe nutzen: das mechanische und chemische Recycling von Abfällen oder auch die energetische Verwertung, bei der Energie gewonnen wird.

Zirkuläre Wertschöpfung bedeutet also, dass neue Produkte, Designs und Verfahren entwickelt werden, bei denen kein sogenannter Abfall mehr entsteht, sondern die Endprodukte nach Gebrauch als Ausgangspunkt einer neuen Wertschöpfung dienen. Das senkt den Rohstoffverbrauch und löst zusätzliche Wachstumseffekte aus.

Um Zirkuläre Wirtschaft messbar, analysierbar und steuerbar zu machen, ist auch die Wirtschaftsstatistik gefordert: Wie können die Kreisläufe dargestellt werden? Wie kann die entstandene Wertschöpfung gemessen werden? Wie sehen die Stoffkreisläufe aus?

160 Netzwerk- und Outputmessung – Indikatorik für transformative Technologiefelder

Muhamed Kudic (2); Katharina Friz (2); Andreas Kladroba (3); Tobias Buchmann (1); Patrick Wolf (1)

1: Zentrum für Sonnenenergie- und Wasserstoff-Forschung Baden-Württemberg (ZSW), Deutschland; 2: Universität Bremen; 3: Stifterverband / FOM Hochschule für Ökonomie & Management
patrick.wolf@zsw-bw.de

SECTION: Business and Market Statistics

Die möglichst präzise Erfassung der FuE-Aktivitäten von Unternehmen stellt eine zentrale Voraussetzung für die Ableitung zielgerichteter und passgenauer innovationspolitischer Maßnahmen sowie für die Konzeption zukunftsweisender Förderlinien dar. Insbesondere der Bedarf an dynamisch angelegten, auf Technologiefeldern basierenden Innovationsstudien steigt hierbei stetig. Den hohen Datenanforderungen solcher Längsschnittstudien werden bestehenden FuE-Zahlenwerke und statistiken jedoch nur bedingt gerecht. Häufig gibt es keine intertemporale Vergleichbarkeit der jeweiligen Grundgesamtheit

und große Schwierigkeiten bei der Kopplung einzelner Erhebungswellen. Zudem erfolgt derzeit keine adäquate Erfassung von Struktur, Dynamik und strategischer Positionierung dieser Unternehmen innerhalb von sich wandelnden Netzwerken. Auch hinsichtlich der Outputmessung decken bisherige Ansätze den Informationsbedarf der Nutzer oft nur unzureichend ab. So sind die bisherigen, meist eher statischen, Outputindikatoren bei sich schnell wandelnden Innovationssystemen nur bedingt aussagekräftig und nur seltenen technologiebasiert.

An diesen Problematiken setzt das vorgestellte Forschungsprojekt an. Übergeordnetes Ziel ist die Gewinnung eines vertieften Verständnisses für das Voranschreiten des Transformationsprozesses des deutschen Innovationssystems sowie seiner technologisch, sektoral und regional abgegrenzten Subsysteme unter genauer Erfassung der Forschungs- und Entwicklungsaktivitäten von Unternehmen und Wissenschaftseinrichtungen. Dabei leisten wir einen wichtigen Beitrag zur Vervollständigung bestehender FuE-Statistiken um sowohl Methoden zur dynamischen Erfassung der Grundgesamtheit und deren Vernetzung als auch neuer/erweiterter Konzepte der Vernetzungsindikatorik und Outputmessung. Unsere Methoden und Verbesserungsansätze veranschaulichen wir hierbei am Beispiel des transformativen, hoch dynamischen Technologiefelds E-Mobilität.

Im bisherigen Projektverlauf wurde eine Methodik zur systematischen, sekundärquellenbasierten Erfassung und Aufbereitung von Kooperationsinformationen entwickelt und auf den Technologiebereich Elektromobilität angewandt. Die erhobenen, aufbereiteten Daten ermöglichen hierbei eine Darstellung des deutschen FuE-Netzes sowie die Durchführung erster Netzwerkanalysen. Gleichzeitig wurde ein Ereignisschema konzipiert, mittels welchem die Historie von Unternehmen nachgezeichnet und das Netzwerk weiter dynamisiert werden soll. Im Rahmen der Entwicklung dynamischer Indikatoren wurde außerdem ein Konzept zur Untersuchung der Plastizität von Netzwerken entwickelt.

161 Wie geht man mit Lücken in Eurostat-Daten um? Vorschläge am Beispiel der FuE-Erhebung.

Andreas Kladroba

Stifterverband für die Deutsche Wissenschaft, Germany
andreas.kladroba@stifterverband.de

SECTION: Business and Market Statistics

Eurostat stellt in seiner Datenbank eine große Menge an Sozial- und Wirtschaftsdaten zur freien Verfügung. Oftmals zeigen sich aber Datenlücken. Meist sind diese in Geheimhaltungsanforderungen begründet. Was kann man tun um dennoch ein vollständiges Bild z.B. von zeitlichen Entwicklungen zu zeichnen? Der Vortrag macht einige einfache Vorschläge dazu. Er geht dabei auf unterschiedliche Szenarien ein, die sich vor allem auf unterschiedliche Informationen beziehen, die sich aus den vorhandenen Daten herauslesen lassen.

162 The Bias of Realized Volatility

Janis Becker; Christian Leschinski

Leibniz Universität Hannover, Germany
becker@statistik.uni-hannover.de

SECTION: Statistics in Finance

Due to its nature as a non-parametric estimate that is consistent for the integrated variance in price processes that behave as semimartingales, realized volatility has become the standard volatility measure and is often even treated as a direct observation of the underlying volatility process. We show based on simple statistical arguments that realized volatility is a biased estimator for the volatility of stock index returns on daily and longer horizons. The bias is negative, so that the stock market risk is systematically underestimated. This effect is demonstrated for a wide range of international stock market indices and the average magnitude of the bias is 14 percent. The RV of the S&P 500, for example, underestimates the mean level of the variance by 15 percent.

Previous contributions on the shortcomings of realized volatility have mostly focused on the effect of market microstructure noise and violations of the assumption that the price process can be observed frictionless at arbitrarily small time intervals. We provide statistical evidence that the bias we find is not due to such market microstructure effects but caused by dependences between aggregates of returns within a trading day, i.e. it is caused by violations of the semimartingale assumption.

163 How to correct the intraday periodicity bias in realized volatility measures

Holger Dette (1); Vasyl Golosnoy (2); Janosch Kellermann (2)

1: Ruhr-Universität Bochum, Fakultät für Mathematik, Deutschland; 2: Ruhr-Universität Bochum, Fakultät für Wirtschaftswissenschaft, Deutschland
janosch.kellermann@rub.de

SECTION: Statistics in Finance

Diurnal fluctuations in volatility are a well-documented stylized fact of intraday price data. We investigate how these intraday patterns (IP) affect both finite sample as well as asymptotic properties of a broad class popular of estimators for integrated volatility (IV) based on functionals of cross-products of intraday returns. We demonstrate that most of the estimators considered in our study exhibit a finite-sample bias due to IP, which can result in invalid statistical inference for IV. We suggest how to construct appropriate correction factors for this bias based on IP estimates. The adequacy of the suggested factors is evaluated both by means of a Monte Carlo study as well as an empirical example.

164 Matrixvariate Factor Model for Realized Covariances

Eugen Ivanov; Yarema Okhrin

Universität Augsburg, Germany
yarema.okhrin@wiwi.uni-augsburg.de

SECTION: Statistics in Finance

Modeling and forecasting covariance matrices of asset returns plays an important role in economic theory and finance. The increasing availability and accessibility of high frequency data shifted the focus of research community to realized measures of covariance, which are intraday counterparts of covariance matrices and were shown to have superior properties. Whereas the univariate case saw a multitude of proposed models, modeling realized measures of covariance lacks truly multivariate methods. One standard approach is to model entries of some matrix decomposition, which leads to loss of interpretability and introduces bias. Another approach considers vectorization of matrix, whereas additional measures have to be taken, in order to ensure positive definiteness. These methods are, however, difficult to estimate and work with in higher dimensions. We propose a matrix variate factor model, which preserves interpretability during estimation and properties of covariance matrix for forecasting. Estimation framework based on Expectation Maximisation algorithm (EM), where factors are estimated with Kalman smoother at each iteration, is presented. We further show the asymptotical consistency of factors estimates under true model parameters when the dimension of observable information is large. An application study presents empirical properties of the proposed model.

165 Big in Japan: Global Volatility Transmission between Assets and Trading Places

Andreas Masuhr

WWU Münster, Germany
andreas.masuhr@wiwi.uni-muenster.de

SECTION: Statistics in Finance

This paper proposes a new framework to model distinct channels of volatility transmission between assets and trading places. The model is estimated using a data set comprising of three stock indices traded at three major trading places: the Nikkei at the Tokyo Stock Exchange, the FTSE at the London Stock Exchange and the S&P500 at the New York Stock Exchange. Strong volatility transmission effects can be observed between London and New York, whereas current volatility in Tokyo mostly depends on past volatility in Tokyo. For the assets in consideration, spillovers are strong across trading zones, but weak across assets, suggesting a close connection between market places but only a loose volatility link between assets. Volatility impulse response functions indicate a long lasting and comparably large response of volatility to shocks in Tokyo, whereas they suggest a quicker volatility decay in London and New York.

166 Regionale Unterschiede in der Verwendung von nicht-informativen ICD-Kodierungen bei kardiovaskulären Todesfällen und deren Auswirkung auf die Vergleichbarkeit von Mortalitätsraten für kardiovaskuläre Erkrankungen

Susanne Stolpe (1,2); Andreas Stang (1,3)

1: Universitätsklinikum Essen, Germany; 2: Leibniz-Universität Hannover, Germany; 3: School of Public Health, Boston, USA

susanne.stolpe@uk-essen.de

SECTION: DGD

Hintergrund:

Die Reliabilität nationaler Mortalitätsdaten wird von der WHO anhand des Anteils nicht-informativer Todesursachen an allen Todesursachen geschätzt. Für Deutschland lag dieser Anteil zwischen 1998 und 2014 im Mittel bei 12% (Finnland 3%, Polen 27%). Die meisten nicht-informativen Todesursachen gehören zu den kardiovaskulären Erkrankungen (I00-I99). Die Verwendung nicht-informativer Todesursachen beeinflusst die Mortalitätsrate kardiovaskulärer Erkrankungen wie z.B. ischämischer Herzerkrankungen. Das Ziel ist es, regionale Unterschiede in der Verwendung nicht-informativer Todesursachen und deren Auswirkungen auf die Mortalitätsraten an ischämischen Herzerkrankungen darzustellen.

Methoden:

Es wurden vom Statistischen Bundesamt (www.gbe-bund.de) alle kardiovaskulären Todesfälle (I00-I99) sowie Todesfälle mit nicht-informativen ICD-10-Kodierungen für 2000 – 2016 geladen. Die nicht-informativen ICD-Kodes wurden gemäß WHO identifiziert. Zur Umkodierung nicht-informativer Kodierungen wurde ein publizierter Algorithmus in vereinfachter Form angewendet.

Ergebnisse:

Der am häufigsten verwendete nicht-informative ICD-Kode bei kardiovaskulären Todesfällen war I50 (Herzinsuffizienz), 2016 die vierthäufigste Todesursache in Deutschland. Es existieren deutliche regionale Unterschiede in Häufigkeit und Art der kodierten nicht-informativen Todesursachen. 2015 betrug der Anteil nicht-informativer Kodes bei kardiovaskulären Todesfällen in NRW 30% (2000: 27%), in Hessen 16% (2000: 23%) und in Sachsen 13% (2000: 13%). Kardiovaskuläre Todesfälle bei Frauen wurden in allen Bundesländern über alle Altersgruppen häufiger mit nicht-informativen ICD-10-Kodes kodiert als bei Männern (insgesamt: 21% vs. 17%). Der Anteil nicht-informativer kardiovaskulärer Todesursachen ist größer bei jüngeren (<55 Jahre) sowie bei sehr alten Verstorbenen (≥ 90 Jahre). Nach Umkodierung von fünf nicht-informativen Todesursachen lag der relative Anstieg der Mortalitätsrate für ischämische Herzerkrankungen 2016 zwischen 20% (Mecklenburg-Vorpommern) und 66% (Bremen).

Diskussion:

Regionale Unterschiede in der Qualität der ausgefüllten Leichenschauscheine und –nachrangig– in der Kodierung der zugrundeliegenden Todesursache beeinflussen erkrankungsspezifische kardiovaskuläre Mortalitätsraten. Der Anteil nicht-informativer Kodes an kardiovaskulären Todesfällen ist hoch und abhängig

von Bundesland, Sterbejahr, Geschlecht und Sterbealter. Insbesondere die Mortalitätsrate für ischämische Herzerkrankungen wird dadurch unterschätzt. Es ist zu überlegen, zum Vergleich der Mortalitätsraten Methoden zur Umkodierung nicht-informativer Todesursachen einzusetzen.

167 Ein Vergleich der Ungleichheit anhand verlorener Lebensjahre zwischen Ost und West Deutschland zwischen 2003 und 2015“

Sebastian Franke

Ludwig-Maximilians-Universität München, Deutschland
sebastian.franke@bwl.lmu.de

SECTION: DGD

Für die Vermittlung der regionalen Mortalität an Kommunen, Behörden und politische Entscheidungsträger bedarf es als ersten Schritt Mortalitätsmaße die zum einen kompakt und zum anderen nicht zu komplex sind. Mortalitätsmaße die sich möglichst in einer (oder wenigen) Zahl(en) zusammenfassen lassen, deren Berechnung jedoch auch für den Laien nachvollziehbar ist.

Verlorene Lebensjahre (YLL), zum Beispiel, sind die Summe an Lebensjahren die einer Gesellschaft aufgrund vorzeitiger Todesfälle verloren gehen. YLLs werden neben der Anzahl der Todesfälle und Sterblichkeitsraten verwendet, da sie das Todesalter berücksichtigen und den jüngeren Altersgruppen dadurch mehr Gewicht verleihen.

In dieser Arbeit verwenden wir das Konzept von YLLs, um die Entwicklung gesundheitlicher Ungleichheiten zwischen Ost- und Westdeutschland, im letzten Jahrzehnt zu analysieren. Hierbei werden neben den 16 Bundesländern als Ganzes, einzelne Bundesländer auf Kreisebene betrachtet.

Ergänzend zu früheren Studien werden neben der kleinräumigeren Darstellung, auch verschiedene Ansätze der verlorenen Lebensjahre sowie deren Standardisierung betrachtet.

Methodik:

Auf der Grundlage von Daten zur altersspezifischen Mortalität und Bevölkerung aus den 16 Bundesländern für den Zeitraum 2003-2015 werden standardisierte und nicht-standardisierte YLLs mittels verschiedene Ansätze für Männer und Frauen sowie der durchschnittliche YLL pro Todesfall für die Gesamtmortalität (ICD10: A00-Y89) und ursachenspezifische Mortalität berechnet. Die Auswahl der Todesursachen aus der Europäischen Kurzliste basiert auf der Häufigkeit der Todesfälle der unter 70-jährigen im Jahr 2014.

Ergebnisse:

Ab 2005 nimmt bei Männern die Ost-West-Ungleichheit bei fast allen wichtigen Krankheitsklassifizierungen wieder zu. Wie bei Männern verlangsamt sich bei Frauen der Rückgang der YLLs im Osten im Vergleich zum Westen, wodurch die Ungleichheit für Männer und Frauen erneut ansteigt.

Fazit:

Die zum Ende der 1990er Jahre verbleibenden gesundheitlichen Ungleichheiten sind seit 2005 wieder gestiegen. Grund hierfür ist die Stagnation bzw. Verlangsamung des Rückgangs der ostdeutschen YLL.

Sowohl für die Betrachtung als auch für den Vergleich von YLL zwischen Regionen oder Jahren ist die Wahl der Standardisierung maßgeblich entscheidend.

168 Das Suizidgeschehen in Deutschland

Holger Leerhoff

Amt für Statistik Berlin-Brandenburg, Deutschland
holger@leerhoff.de

SECTION: DGD

Der Konferenzbeitrag befasst sich mit einer Betrachtung des Suizidgeschehens in der Bundesrepublik. Rund 10.000 Menschen nehmen sich in Deutschland jährlich das Leben. Umfassende und belastbare statistische Daten zu diesem Phänomen sind ein wichtiger Baustein für eine effektive Suizidprävention. Auf Grundlage der amtlichen Todesursachenstatistik werden in einem laufenden Projekt die entsprechenden Todesfälle (ICD 10: X60-X84) seit 1992 hinsichtlich unterschiedlicher soziodemographischer und weiterer Merkmale analysiert und in diesem Vortrag erste Ergebnisse vorgestellt.

Neben einer längsschnittlichen Betrachtung der Suizidrate im regionalen und internationalen Vergleich wird mittels eines für diese Analysen erstellten synthetischen Jahrgangs der Frage nachgegangen, inwiefern bei den Todesfällen soziodemographische Merkmale der Verstorbenen (Geschlecht, Alter, Familienstand, Konfession, Wohnort) mit Umfang und den individuellen Umständen des Suizids – insbesondere dem Sterbedatum, aber auch Art und Ort der Selbsttötung – in Zusammenhang stehen. Ein Fokus wird dabei auf einer Quantifizierung der oft berichteten Häufung der Suizide um Weihnachten und den Geburtstag der Verstorbenen liegen. Es soll zudem versucht werden, Zusammenhänge zwischen der Anzahl der Suizide und markanten Ereignissen wie der Wirtschaftskrise und sportlichen Großereignissen aufzuzeigen. Exemplarisch für Berlin soll auch einer mögliche Korrelation zwischen dem Wetter und dem Suizidgeschehen nachgegangen werden.

169 Die wirtschaftliche Bedeutung von Hochschulen für ihre Region

Britta Stöver

Institut für Statistik an der Leibniz Universität Hannover, Germany
stoever@statistik.uni-hannover.de

SECTION: Regional Statistics

Hochschulen haben einen nicht zu vernachlässigenden Einfluss auf Angebot und Nachfrage und re-präsentieren damit einen wichtigen wirtschaftlichen Akteur in ihrer Region. Mit diesem Beitrag wird auf Basis einer Multiplikatoranalyse für die Nachfragewirkungen sowie eines räumlich-ökonomischen Schätzmodells (spatial panel) für die Angebotswirkungen die Bedeutung der Hochschulstandorte in Niedersachsen auf Kreisebene quantifiziert, verglichen und klassifiziert. Im Vergleich zu zahlreichen anderen Beiträgen, welche die ökonomische Bewertung von Hochschulstandorten vornehmen, stellt dieser Beitrag nicht nur eine bestimmte Hochschule bzw. Hochschulregion losgelöst von anderen Standorten in den Fokus, sondern bietet ein umfassendes, vollständiges und konsistentes Bild aller Hochschulstandorte in Niedersachsen, sodass auch ein Vergleich der Standorte untereinander möglich ist. Die einkommensinduzierten direkten und indirekten Nachfrageeffekte werden auf Basis eines detaillierten Datensatzes der Hochschulstatistik in Kombination mit Einkommens- und Beschäftigungsmultiplikatoren aus einer regionalen Input-Output-Tabelle berechnet. Die angebotsseitigen Effekte, also die Wirkungen, die aus der Lehre und Forschung entstehen, werden auf Basis der Humankapitaltheorie und der Theorie zur Wissensverbreitung (knowledge spillover theory) mit Hilfe von spatial panel regressionen geschätzt. Die Ergebnisse vermitteln einen vollständigen und nachvollziehbaren Eindruck über die wirtschaftliche Bedeutung jedes einzelnen Hochschulstandorts in Niedersachsen und ermöglichen Vergleiche zwischen den Standorten sowie Einordnungen bzw. Gruppierungen.

170 Daten zur Identifikation ethnisch diversifizierter Sozialräume

Guido Oemmelen

microm, Germany
g.oemmelen@microm.de

SECTION: Regional Statistics

Die ethnische Herkunft wird häufig über die Sprachabstammung (Onomastik) von Personen ermittelt. Dafür werden Vor- und Nachname hinsichtlich der sprachlichen Herkunft analysiert. Die Zuordnung basiert auf Listen, in denen die sprachliche Abstammung von Namen verzeichnet ist, ergänzt durch internationale Namensverzeichnisse. Dabei wird der Migrationshintergrund bewusst von der Nationalität unterschieden, da der ethnische und kulturelle Hintergrund entscheidend die Werteorientierung und das Konsumverhalten dieser Menschen beeinflusst.

Die Basis für die Übertragung in den Raum bildet neuerdings eine aktuelle, ständig gepflegte Adress-datei von ca. 70 Mio. Personen in Deutschland, im Besitz eines namhaften Adressdienstleisters.

Die prozentualen Anteile zehn verschiedener ethnischer Gruppen werden datenschutzkonform aggregiert und ab der kleinräumigen Gliederung der microm PLZ8-Gebiete flächendeckend für die BRD ausgewiesen und analysiert.

Der Vortrag stellt die Übertragung der Daten aus einem nicht-amtlichen Quasi-Register dar und zeigt auf, welche Möglichkeiten bestehen auf dieser Datenbasis regionale Vergleiche anzustellen und ethnisch diversifizierte Sozialräume zu identifizieren.

171 Daseinsvorsorge und interkommunale Kooperation in kleineren Städten und Gemeinden

Madeline Kaupert

Bundesinstitut für Bau-, Stadt- und Raumforschung, Germany
madeline.kaupert@bbr.bund.de

SECTION: Regional Statistics

Derzeit steht die Forderung nach „gleichwertigen Lebensverhältnissen“ auf der gesellschaftlichen und politischen Tagesordnung. Der demographische Wandel betrifft zwar das ganze Bundesgebiet, da das Durchschnittsalter der Bevölkerung überall steigt – allerdings sind die Folgen des demographischen Wandels für Gebiete außerhalb der Großstadregionen, die zusätzlich von Wanderungsverlusten und wirtschaftlichem Strukturwandel betroffen sind, gravierender. Hier stellt sich mancherorts die Frage, wie und in welchem Umfang öffentliche Infrastruktur erhalten bleiben kann. Die Forderung nach gleichwertigen Lebensverhältnissen ist keineswegs neu, sondern tritt auf ein Netz bereits bestehender Ansätze zur Förderung strukturschwacher Regionen. Seit 2010 gehört dazu auch das Programm „Kleinere Städte und Gemeinden – überörtliche Zusammenarbeit und Netzwerke“ aus der Familie der Städtebauförderprogramme.

Die Begleitinformationen und Monitoringdaten der Städtebauförderung sind Grundlage für die Auswertung des Programms. Aus der kombinierten Analyse mit Daten der laufenden Raumb Beobachtung lassen sich vertiefte Erkenntnisse über kleinere Kommunen gewinnen. Die Daten zeigen, in welchen Bereichen der Daseinsvorsorge die Kommunen selbst die drängendsten Probleme ausmachen und wo sie aktiv werden. Außerdem können auf dieser Grundlage Rückschlüsse auf die aktuelle Förderung gleichwertiger Lebensverhältnisse gezogen werden. Beispielsweise profitiert der ländliche Raum gemessen an seiner Bevölkerung bereits heute überproportional von den Mitteln der Städtebauförderung.

Eine Besonderheit im Programm „kleinere Städte und Gemeinden“ ist die vorrangige Förderung interkommunaler Kooperationen. Union und SPD haben im Koalitionsvertrag vereinbart, interkommunale Kooperationen zu stärken. Die BBSR-Datenbank zum Programm enthält Informationen über mehr als 200 Zusammenschlüsse unabhängiger Kommunen. Sie ist damit eine seltene statistische Quelle für eine kommunale Strategie, die bisher normalerweise nur in Fallstudien untersucht wird.

172 Selecting variables via clustering within regularized prediction in high dimensions

Rainer von Sachs (1); Marion Rebecca (1); Govaerts Bernadette (1); Lederer Johannes (2)

1: Université catholique de Louvain, Belgium; 2: Ruhr-Universität Bochum, Germany
rvs@uclouvain.be

SECTION: Statistic in Natural Science and Technology

This presentation addresses the problem of selecting important, potentially overlapping groups of predictor variables in linear models such that the resulting model satisfies a balance between interpretability and prediction performance.

This problem is motivated by data from the field of chemometrics (such as Nuclear Magnetic Resonance or Near-Infrared spectra) where, due to correlation between predictors from different groups (i.e. variable group „overlap“), identifying groups during model estimation is particularly challenging.

Our proposition for addressing the variable clustering problem is to incorporate prior information about data structure into existing clustering methods using a „localisation by weighting“ approach. This clustering approach would then be integrated into regularising prediction methods, such as the Cluster Elastic Net by Witten et al (2014). In this talk, after giving a gentle introduction into the problem and its state of the art, we briefly discuss different approaches of variable clustering with respect to their ability to handle the overlapping group nature. In particular, we concentrate on our own proposal, which is Weighted Clustering around Latent Variables (WCLV), motivated by the work of Vigneau et al (2003, 2015).

References:

Vigneau, E. and Qannari, E. (2003). Clustering of variables around latent components. *Communications in Statistics – Simulation and Computation*, 32(4):1131–1150.

Vigneau, E., Chen, M., and Qannari, E. M. (2015). Clustvarlv: An R package for the clustering of variables around latent variables. *R Journal*, 7(2).

Witten, D., Shojaie, A., and Zhang, F. (2014). The Cluster Elastic Net for high-dimensional regression with unknown variable grouping. *Technometrics*, 56, 112-122.

173 Asymptotics for High-Dimensional Covariance Matrices of Factor Models

Monika Bours ; Ansgar Steland

RWTH Aachen University, Deutschland
bours@stochastik.rwth-aachen.de

SECTION: Statistic in Natural Science and Technology

In this talk we consider high dimensional vector time series with a factor model structure. It is assumed that the number of factors is allowed to go to infinity.

Factor models are widely used in various areas including psychometrics, marketing, finance as well as natural sciences and technology. For example, factor models are a key tool for financial risk analysis and macroeconomic forecasting of indicators such as the GDP (Gross Domestic Product) and inflation.

Since large datasets are becoming increasingly available in many disciplines the analysis of high dimensional time series has become an highly active area. The estimation of high-dimensional variance-covariance matrices is of particular interest, but often only an intermediate step, since interest focuses on the behavior of functions of the sample variance-covariance matrix, especially bilinear forms which naturally arise when studying projection type statistics.

We establish new results on distributional approximations for bilinear functions of sample variance-covariance matrices in terms of Brownian motions. In the high dimensional setting, where also the dimension is allowed to go to infinity, these approximations by Gaussian processes hold true without any constraints on the dimension, the sample size or their ratio. Our results are valid for uniformly bounded projection vectors in the l_1 norm which arise either naturally or by construction in many statistical problems like change-point analysis, sparse financial portfolio selection and shrinkage estimation.

A simulation study illustrates the performance of this theory. For example we will use our results to detect changes in the variance of a projection.

174 Detecting Changes in the Second Moment Structure of High-Dimensional Sensor-Type Data in a K-Sample Setting

Nils Mause; Ansgar Steland

RWTH Aachen University, Deutschland
mause@stochastik.rwth-aachen.de

SECTION: Statistic in Natural Science and Technology

In this talk we consider multiple samples of high-dimensional vector time series generated at independent locations of sensors within a certain time interval. These sensors may collect and transmit data at different sampling frequencies, such that the resulting sample sizes may also be different. Projections based on such high-dimensional time series appear naturally in many statistical procedures and applications, such as the Principal Component Analysis and Portfolio Optimization, and are a common method of dealing with high-dimensional data sets.

Within the high-dimensional framework where not only the time horizon shall go to infinity but also the dimension of the data is allowed to grow with the time horizon we establish new approximation results for a sum of squared bilinear forms based on the sample variance – covariance matrices as well as for bilinear forms based on the pooled sample variance – covariance matrix. These approximation results do not depend on any constraint on the ratio of dimension and sample size and further allow us to deal with a change in variance problem, where, under the null hypothesis of no change, the data is supposed to form a stationary vector time series with mean zero and a variance – covariance matrix which can either be known or unknown. We construct CUSUM based change – point test statistics and derive their asymptotic behaviour for each of these situations. Simulations and an application on a data set generated by pressure sensors are used to illustrate the finite sample performance of our change-point tests.

This is joint work with Ansgar Steland from the Institute of Statistics, RWTH Aachen University, Germany. It was supported by the German Research Foundation / Deutsche Forschungsgemeinschaft (DFG) [grant number STE 1034/11-1].

175 BA-Statistik für Datenjournalismus nutzen

Thomas Weißbrodt; Stefan Werth

Statistik der Bundesagentur für Arbeit, Germany
thomas.weissbrodt@arbeitsagentur.de

SECTION: Data Journalism

Die Statistik der Bundesagentur für Arbeit erstellt und veröffentlicht umfassende Daten über den Arbeitsmarkt. Hinter den Statistiken der Bundesagentur für Arbeit verbergen sich Stories. Neben Berichtsformaten wie Tabellen und Reports bietet die Statistik der BA interaktive Anwendungen an. Mit diesen können schnell und einfach Visualisierungen erstellt werden, die z.B. in Vorträgen, Pressemitteilungen oder wissenschaftlichen Publikationen verwendet werden können. Erfahren Sie in diesem Vortrag, wie Sie die Angebote der Statistik der BA für Datenjournalismus nutzen können.

Die leichte Verfügbarkeit der Visualisierungen darf über eine Sache nicht hinwegtäuschen: Hinter diesen Darstellungsformen verbergen sich komplexe statistische und rechtliche Konstrukte und damit die Gefahr von Fehlinterpretationen. Um die visualisierten Daten verstehen und mit ihrer Hilfe Arbeitsmarktfragen beantworten zu können, ist die Auseinandersetzung mit den Inhalten erforderlich – eine zu erbringende Übersetzungsleistung der Datennutzer, die statistische Informationen weitergeben.

176 Entwicklung von Methoden und Tools für eine datengestützte Wissenschaftskommunikation

Lars Koppers (1); Sarah Kohler (1); Meik Bittkowski (2); Nikolai Promies, (1); Volker Stollorz (2); Markus Lehmkuhl (1)

1: Karlsruher Institut für Technologie, Germany; 2: Science Media Center Germany
lars.koppers@kit.edu

SECTION: Data Journalism

Üblicherweise beobachten Wissenschaftsjournalist*innen nur einen kleinen Teil an Top-Journalen, um über dort veröffentlichte Studien zu berichten. Im Zuge der Digitalisierung ist es notwendig, nicht nur diese Top-Journale, sondern auch weitere Quellen, wie zum Beispiel Preprint-Server in den Blick zu nehmen, um die Beobachtungs- und Bewertungsfunktion des Wissenschaftsjournalismus auch in Zukunft wahrnehmen zu können. Das „Science Media Center Germany“ entwickelt im „SMC lab“ Software und Services für die eigene Redaktion und für die journalistische Community. In diesem Beitrag werden zwei Projekte vorgestellt.

In einem Kooperationsprojekt mit dem KIT wird ein Monitoring-System für den Preprint-Server bioRxiv entwickelt. Anhand von Metadaten wie Abrufstatistiken und bibliometrischen Kennzahlen wie dem „Altmetric Attention Score“ sollen neue Paper identifiziert und empfohlen werden, die im jeweiligen Fachgebiet bereits stark diskutiert werden. Mit Hilfe von Textmining-Ansätzen wie der latent dirichlet allocation wird zudem eine thematische Auswahl ermöglicht, je nach Suchstrategie der Wissenschaftsredaktion.

Solch automatisierte Tools ermöglichen es Wissenschaftsjournalist*innen trotz ihrer begrenzten Zeit verschiedene wissenschaftliche Fachgebiete zu überblicken und über aktuelle Entwicklungen in der Wissenschaft zu berichten.

In dem Projekt „Operation Explorer“ werden Daten zu Diagnosen, Operationen und Prozeduren in Krankenhäusern gesammelt und für Journalist*innen aufbereitet. Die Ergebnisse werden auf Landkreis-Ebene aggregiert und auf Deutschlandkarten visualisiert. Dabei treten die üblichen Probleme der räumlichen Statistik auf.

177 Quantifizierung der Stabilität der Latent Dirichlet Allocation mithilfe von Clustering auf wiederholten Durchläufen

Jonas Rieger; Carsten Jentsch; Jörg Rahnenführer

Technische Universität Dortmund, Deutschland
jonas.rieger@tu-dortmund.de

SECTION: Data Journalism

Die Modellierung von Themenstrukturen in Texten (Topic Modelle) bietet Möglichkeiten zur Darstellung großer Textkorpora. Eine weit verbreitete Methode ist die Latent Dirichlet Allocation (LDA), ein generatives probabilistisches Modell, das einzelne Texte einer Textsammlung als Mischungen von latenten Themen modelliert. Jedes Thema zeichnet sich durch seine Wortverteilung aus. Die konkreten Zuordnungen von Wörtern zu Themen unterliegen dabei dem Zufall: zum einen sind die initialen Themenzuordnungen zufällig; zum anderen werden diese in der Modellierungsprozedur basierend auf allen anderen Zuordnungen mithilfe von Gibbs Sampling stochastisch neu zugeordnet. Aus diesem Grund führen wiederholte LDA-Läufe auf denselben Textdaten zu unterschiedlichen Ergebnissen. Dieser Fakt wird in der Praxis häufig vernachlässigt.

Wir untersuchen die Stabilität des Verfahrens LDA durch den Vergleich wiederholter Durchläufe der Zuordnungen von Wörtern zu Themen. Die dafür benötigten Ähnlichkeiten von Themen werden durch einen modifizierten Jaccard Koeffizienten bestimmt. Unser neuer Algorithmus, der auf der Idee basiert, dass zwei Modelle eine Reihe von Paaren ähnlicher Themen erzeugen, führt zu einem Maß für die Quantifizierung der Stabilität von LDA-Läufen.

Unsere Ergebnisse zeigen, dass dieses Ähnlichkeitsmaß für die Beurteilung der Stabilität von LDA-Läufen nützlich ist. Neben LDAs lassen sich so auch Ergebnisse anderer Topic Modelle, die ihre Themen durch Wortverteilungen charakterisieren, vergleichen. Zudem stellen wir eine Methode zur Verbesserung der Reliabilität eines repräsentativen LDA-Durchlaufs vor, um die Reproduzierbarkeit der Ergebnisse von Topic Modellen zu verbessern. Diese Methode basiert auf der Idee eine Vielzahl von Durchläufen durch einen besten Repräsentanten zu beschreiben. Dieser zentrierte LDA-Durchlauf zeichnet sich durch die empirisch größte Ähnlichkeit zu allen berechneten Durchläufen gemäß unseres neuen Ähnlichkeitsmaßes aus.

Das vorgestellte Maß wird in Zukunft für den Vergleich von LDA-Läufen, die auf unterschiedlichen Textdaten basieren, verallgemeinert werden. In diesem Zusammenhang können Vergleiche verschiedener Medien mit gleicher Publikationsform (z.B. verschiedene Zeitungsredaktionen) oder auch Vergleiche verschiedener Publikationsformen (z.B. Print vs. Twitter) des gleichen Mediums analysiert werden.

178 Bernstein-von Mises Theorems for Compound Poisson Processes

Richard Nickl (1); Jakob Söhl (2)

1: University of Cambridge, UK; 2: Delft University of Technology, The Netherlands
j.soehl@tudelft.nl

SECTION: Bayesian Analysis in Natural and Social Science

We study nonparametric Bayesian statistical inference for the parameters governing a compound Poisson process. A compound Poisson process is a pure jump process where the jump times follow a Poisson process with intensity λ and the jump sizes are i.i.d. draws from a jump measure μ . Compound Poisson processes are characterized by their Lévy measure ν which is the product of their intensity λ and their jump measure μ . A high-dimensional wavelet series prior for the Lévy measure ν is devised and the posterior distribution arises from observing discrete samples with fixed observation distance, giving rise to a nonlinear inverse inference problem. We derive contraction rates in uniform norm for the posterior distribution around the true Lévy density that are optimal up to logarithmic factors over Hölder classes, as sample size n increases. We prove a functional Bernstein-von Mises theorem for the distribution functions of both μ and ν , as well as for the intensity λ , establishing the fact that the posterior distribution is approximated by an infinite-dimensional Gaussian measure whose covariance structure is shown to attain the Cramér-Rao lower bound for this inverse problem. As a consequence posterior based inferences, such as nonparametric credible sets, are asymptotically valid and optimal from a frequentist point of view.

179 Bayesian Inference of the Multi-Period Optimal Portfolio for an Exponential Utility

Nestor Parolya (1); Taras Bodnar (2); Wolfgang Schmid (3); David Bauder (4)

1: Delft University of Technology, Netherlands; 2: Stockholm University, Sweden; 3: Viadrina University, Germany; 4: cellent GmbH, Dresden
n.parolya@tudelft.nl

SECTION: Bayesian Analysis in Natural and Social Science

We consider the estimation of the multi-period optimal portfolio obtained by maximizing an exponential utility. Employing Jeffreys' non-informative prior and the conjugate informative prior, we derive stochastic representations for the optimal portfolio weights at each time point of portfolio reallocation. This provides a direct access not only to the posterior distribution of the portfolio weights but also to their point estimates together with uncertainties and their asymptotic distributions. Furthermore, we present the posterior predictive distribution for the investor's wealth at each time point of the investment period in terms of a stochastic representation for the future wealth realization. This in turn makes it possible to use quantile-based risk measures or to calculate the probability of default. We apply the suggested Bayesian approach to assess the uncertainty in the multi-period optimal portfolio by considering assets from the FTSE 100 in the weeks after the British referendum to leave the European Union. The behaviour

of the novel portfolio estimation method in a precarious market situation is illustrated by calculating the predictive wealth, the risk associated with the holding portfolio, and the default probability in each period.

180 Bayesian mean-variance analysis: Optimal portfolio selection under parameter uncertainty

Taras Bodnar (1); David Bauder (2); Nestor Parolya (3); Wolfgang Schmid (4)

1: Stockholm University, Sweden; 2: cellent GmbH, Dresden, Germany; 3: Delft University of Technology, Netherland; 4: European University Viadrina, Germany
taras.bodnar@math.su.se

SECTION: Bayesian Analysis in Natural and Social Science

The paper solves the problem of optimal portfolio choice when the parameters of the asset returns distribution, like the mean vector and the covariance matrix are unknown and have to be estimated by using historical data of the asset returns. The new approach employs the Bayesian posterior predictive distribution which is the distribution of the future realization of the asset returns given the observable sample. The parameters of the posterior predictive distributions are functions of the observed data values and, consequently, the solution of the optimization problem is expressed in terms of data only and does not depend on unknown quantities. In contrast, the optimization problem of the traditional approach is based on unknown quantities which are estimated in the second step leading to a suboptimal solution.

We also derive a very useful stochastic representation of the posterior predictive distribution whose application leads not only to the solution of the considered optimization problem, but provides the posterior predictive distribution of the optimal portfolio return used to construct a prediction interval. A Bayesian efficient frontier, a set of optimal portfolios obtained by employing the posterior predictive distribution, is constructed as well. Theoretically and using real data we show that the Bayesian efficient frontier outperforms the sample efficient frontier, a common estimator of the set of optimal portfolios known to be overoptimistic.

181 Bias correction for non-probability samples

Monica Pratesi (1); Luciana Quattrocioch (2); Gaia Bertarelli (1); Alessandro Gemignani (3); Caterina Giusti (1)

1: University of Pisa, Italien; 2: Istat; 3: Interuniversity Research Center Camilo Dagum
monica.pratesi@unipi.it

SECTION: Joint DStatG & SIS – Survey Statistics

Considering the 2030 Agenda for Sustainable Development, SDG 4 aims to ensure inclusive and equitable quality education and promote lifelong learning opportunities for all. Educational Poverty (EP) is defined as deprivation of the ability to learn, experiment, develop and freely flourish skills, talents and aspirations. It means being excluded from acquiring the skills needed to live in a world characterized by knowledge-based economy, rapidity and innovation. EP is a latent trait, only indirectly measurable through a collection of observable variables and indicators purposively selected as micro-aspects, contributing to the latent macro-dimension. In the last few years, the Italian National Statistical Institute (ISTAT) started to measure it by a multidimensional index, the Educational Poverty Index (EPI). A problem with this index is that it is based on survey direct estimates, which are reliable only at regional (NUTS 2) level, while to intervene on the phenomenon it is important to obtain information at a finer geographical level. This implies the need of Small Area Estimation (SAE) methods. In this paper SAE models are applied to the indicators selected to measure the EP dimensions of the multidimensional EPI proposed by ISTAT. The aim is to focus on the multidimensional aspect of EP and to enhance the knowledge of the spatial distribution of EP at local level in Italy, in order to help the policy maker to address resources towards the areas where the phenomenon is strongly present.

182 Bias correction for non-probability samples

Jan Pablo Burgard

Universität Trier, Deutschland
burgardj@uni-trier.de

SECTION: Joint DStatG & SIS – Survey Statistics

Small area estimation encompasses techniques to obtain population estimates at a low aggregation level albeit the sample information is scarce. One of the major concerns is the issue of measurement errors, which typically are more severe in small sample size settings. In this talk, different approaches to perform small area estimation under measurement errors are shown. The treated measurement errors can be either normally distributed, or only the first and second moment are known or the distribution is completely unknown. Further, it is also discussed how to obtain mean squared error estimates under the different scenarios.

183 Volatility spillover along the supply chains: A network analysis on economic links

Theo Berger

University of Bremen, Germany
thberger@uni-bremen.de

SECTION: Statistics in Finance

We introduce a financial network approach to quantify the impact of counterparty risk on firms' daily market risk, measured via conditional volatility. Translating conditional volatility into a Value-at-Risk (VaR) framework allows us to identify extreme losses beyond an estimated loss limit and to determine volatile market regimes. We find that suppliers are exposed to additional fundamental risk that is not captured by the suppliers' market beta, which gets transferred along the supply chains. The identified risk spillover impact both dimension and quality of the suppliers' market risk assessment: If customers experience large losses beyond their individual VaR limit, suppliers' variance forecasts increase by (up to) 1% and the probability of suppliers' extreme losses doubles the next day.

184 Conditional skewness and kurtosis of aggregated normal mixture GARCH returns with application to Value-at-Risk and Expected Shortfall forecasting

Markus Haas

Universität Kiel, Germany
markus.haas@qber.uni-kiel.de

SECTION: Statistics in Finance

The first four conditional moments of aggregated returns generated by mixed normal GARCH (MixN-GARCH) processes are derived. In the MixN-GARCH model, the conditional distribution is a discrete Gaussian mixture with each component variance being endowed with its own conditional volatility process. The result is a highly flexible yet still very tractable structure incorporating in a most natural way time-varying conditional skewness and kurtosis, which are important characteristics of many financial time series and relevant for accurate risk assessment. The MixN-GARCH process often provides a plausible disaggregation of the stochastic mechanism generating returns, and it has been shown to deliver excellent out-of-sample density and risk measure forecasts. Besides illustrating the considerable flexibility and great variety of the conditional density profiles that can emerge from this class of conditional volatility models, the results turn out to be useful for approximating multi-step-ahead Value-at-Risk and Expected Shortfall measures of aggregated returns.

185 Asymptotic Properties of GARCH Models in the Presence of Time Varying Unconditional Variance. A Simulation Study

Oliver Old

FernUniversität in Hagen, Germany
oliver.old@fernuni-hagen.de

SECTION: Statistics in Finance

In this paper the asymptotic and finite sample properties of symmetric GARCH and asymmetric GJR-GARCH models in presence and in absence of time varying long term variance are considered. In particular, the semi-parametric spline-GARCH model is investigated by Monte-Carlo simulation, where the true parameter values are taken from estimated real equity index data. As a proxy for the behaviour of equity indices of developed countries, the S&P500 Index is estimated with the Quasi-Maximum-Likelihood (QML) method for different conditional heteroscedastic models (GARCH, GJR-GARCH, spline-GARCH and spline-GJR-GARCH). The estimated S&P500 parameter values are used to simulate a broad range of 6 different time-series lengths 100, 500, 1000, 5000, 10000, 25000 and 4 different numbers of spline knots 1,4,9,14, combining to a total amount of 60 different model setups. To the best of my knowledge, there exist only a few limited simulation studies that focus on the spline-GARCH model. The main contribution of this paper is therefore to highlight the behaviour of the QML estimates, when the long-term variance is implemented by the spline-GARCH model. Beside this, the paper provides a least-square approach to get useful starting values for the demanding numerical estimation routine. Furthermore analytical derivatives for the spline-GARCH model and some testing procedures for the QML estimates are provided.

186 New Approaches in Testing High-Dimensional EU Portfolios

Taras Bodnar (1); Solomiia Dmytriv (2); Yarema Okhrin (3); Nestor Parolya (4); Wolfgang Schmid (2)

1: Stockholm University, Sweden; 2: European University Viadrina, Germany; 3: University of Augsburg, Germany; 4: Delft University of Technology, Netherlands
dmytriv@europa-uni.de

SECTION: Statistics in Finance

The subject of this talk is the weights of the expected utility (EU) portfolio. Our primary focus lies on a shrinkage-based approach for estimating optimal EU portfolio weights. We propose a test for the EU portfolio weights based on the estimated shrinkage intensity and discuss the asymptotic behavior of the proposed test statistic under a high-dimensional asymptotic regime. We provide an extensive simulation study where the test performance was measured in terms of power functions and receiver operating characteristic curves. The theoretical findings are implemented in an empirical application to the returns of the S&P 500.

187 Statistical analysis of joint pausing in parallel spike trains

Matthias Gärtner; Sevil Duvarci; Jochen Roeper; **Gaby Schneider**

Goethe University Frankfurt, Germany
schneider@math.uni-frankfurt.de

SECTION: Statistic in Natural Science and Technology

Periods with surprisingly few events, so-called „pauses“, have gained increasing attention in the analysis of parallel spike trains. In particular, simultaneous pausing may have an important impact on information processing. Available analysis methods that first identify pauses in individual spike trains are often threshold dependent and can fail to identify joint pauses in parallel processes that are visible to the eye.

We therefore introduce techniques that rely on the joint analysis of parallel spike trains. A statistic called pausiness is introduced that works with the superposition of two processes. This statistic measures the degree of synchronous pausing in spike train pairs and avoids threshold-dependent identification of individual pauses. A graphic termed the „cross-pauseogram“. compares the joint pausiness of two spike trains with its time shifted analogue and can thus indicate simultaneous pausing. When assessing statistical significance, we use a stochastic point process model that also comprises synchronous spiking events. Parameter estimates are obtained from auto- and cross-correlation functions.

The techniques are applied to a data set of dopaminergic neurons recorded from freely moving mice, where joint pausiness was observed in about 20% of the investigated neuron pairs.

188 Improving the ARL profile and the accuracy of its calculation for Poisson EWMA charts

Sven Knoth

Helmut Schmidt University Hamburg, Germany
knoth@hsu-hh.de

SECTION: Statistic in Natural Science and Technology

The Poisson exponentially weighted moving average (PEWMA) chart was proposed by Borrór, Champ and Rigdon (1998) to monitor the mean of counts of nonconformities. For small in-control values of this mean, the Poisson distribution is skewed. Hence, while setting up the control limits one should take care of this skewness – limits symmetric around the the in-control mean induce often unsatisfactory ARL profiles. We offer a nice way to handle the problem. Moreover, dealing with count data makes matching pre-defined in-control ARL levels difficult. Here we illustrate that this is a minor problem for EWMA charts (in contrast to CUSUM charts). Eventually, we provide an improved Markov chain technique to calculate the ARL. In summary, setting up suitable two-sided EWMA designs for monitoring the mean of Poisson data is now a straightforward task (functions are implemented in R).

BORROR, CHAMP, RIGDON (1998). „Poisson EWMA control charts“. *Journal of Quality Technology*, 30(4), pp. 352-361.

189 Time series diagnostics and forecasting of rainfall data

Markus Fritsch (1); Harry Haupt (1); Joachim Schnurbus (1); Philipp Sibbertsen (2); Kai Rouven Wenger (2)

1: University of Passau, Germany; 2: Leibniz University Hannover, Germany
wenger@statistik.uni-hannover.de

SECTION: Statistic in Natural Science and Technology

The characteristics of rainfall time series data, such as seasonal, memory and break patterns, and corresponding statistics depend on the level of aggregation, and hence may require different modeling and forecasting strategies.

We contribute to the literature on modeling rainfall data by providing a systematic analysis of long daily time series for 30 cities, classified according to a climate zone classification: First, we employ time series diagnostics on different aggregation levels and check their respective sensitivity. In particular we compare our findings to the existing literature investigating the memory properties of rainfall processes using recent state-of-the-art methods. For daily time series we find indications for long memory, while for monthly time series such indications mostly disappear, in line with results reported in the literature. Second, the first-step diagnostics allow us to formulate a selection of appropriate candidate forecasting approaches. Using our data sets we check the properties of all candidate models and compare their prediction performance using selected criteria and forecasting horizons.

Our results suggest that although the aggregation level mainly depends on the question of interest, statistical analysis of different aggregation levels may provide additional insights and improved forecasts.

190 Monitoring image processes assuming spatially correlated pixels.

Ivan Semeniuk (1); Wolfgang Schmid (1); Yarema Okhrin (2)

1: Europa University Viadrina, Germany; 2: Augsburg University, Germany
semeniuk@europa-uni.de

SECTION: Statistic in Natural Science and Technology

In recent years, we observe extreme changes how quality features of manufactured products are designed and inspected. New measurement tools, i.e., fast multi-stream high-speed sensors, are invented and used in emerging technologies, e.g., additive manufacturing. The modeling and monitoring problems obtained by new inspection methods are quite complex. For such type of processes completely different types of quality characteristics must be monitored – mainly not the mean, the variance, the covariance matrix or a simple profile, but, e.g., the shape, surfaces and images, etc. As a result many new approaches have been developed in the fields of image analysis, spatial statistics and spatio-temporal modeling. A huge number of tools is available to model the underlying processes. For that reason, Statistical Process Control faces new problems. The main issue of this paper lies in the development of monitoring schemes for such processes. We concentrate on monitoring image processes. Since even a typical image of size 1920x1080 consists of more than 2 million pixels, we encounter a high-dimensional data set. Therefore, techniques for the dimension reduction must be applied while monitoring. Previous research in the field

of image monitoring mainly relies on the assumption that the pixels within an image are uncorrelated. However, this does not hold in many cases, e.g., quantum-limited imaging, such as in X-ray and nuclear-medicine imaging. In our work, new procedures for monitoring image processes with correlated pixels are presented. They are based on the generalized likelihood method and multivariate exponential smoothing taking into account the local correlation structure. Within an extensive simulation study we assess the performance of the suggested methods and compare them to existing approaches.

191 Small area poverty indicators adjusted using local price indexes

Monica Pratesi (1); Gaia Bertarelli (1); Luigi Biggeri (2); Caterina Giusti (1); Stefano Marchetti (1); Francesco Schirripa-Spagnolo (1)

1: University of Pisa, Italien; 2: University of Florence, Italien
monica.pratesi@unipi.it

SECTION: MAKSWELL EU

Policy makers and stakeholders require reliable measures and indicators in order to design, implement and evaluate intervention policies effectively both at national and local level. Usually, to obtain reliable estimates at local level require to resort to small area estimation techniques. In this work we focus on estimating the incidence and intensity of monetary poverty at sub-regional level taking into account the different price level within the country. Indeed, in Italy the North- South divide has turned off into a strong different price level, which can affect the poverty threshold. The local price level is accounted by purchasing power parity indexes computed at sub-regional level from two different big data sources: 1. from retail scan data on regional and sub-regional retail volumes (units) and price for food and grocery (Istat/Nielsen), and 2. the transaction prices of houses ade available by the Revenue Agency at Italian sub-municipality level Revenue Agency – OMI). Sub-regional level poverty estimates will be obtained using area level small area models, which link direct unreliable estimates to aggregated auxiliary information, often easily available. This work is motivated by previous evidence of the influence of sub-national regional) price indexes on local poverty threshold, depicting a different poverty scenario then that obtained without considering them. This work is partially supported by the MAKSWELL European project G.A. no. 770643.

192 Alternative spatial data sources for small area estimation in developing countries

Paul Smith; Nikos Tzavidis; Angela Luna; Jessica Steele; Kristine Nilsen

University of Southampton
P.A.Smith@soton.ac.uk

SECTION: MAKSWELL EU

Monitoring of socio-demographic indicators at subnational levels poses a challenge for most countries in the world due to the insufficient sample sizes of the surveys containing the main variables of interest. In developed countries this issue has been successfully addressed by using small area estimation (SAE) methods, which improve upon the estimates that can be directly obtained from survey data only by „borrowing strength“ from hard sources such as censuses and administrative registers. For developing countries on the other hand, and particularly for low-income ones, extensive use of SAE methods has been hindered by the lack of up-to-date high quality hard sources. Spatial big data, which are generally available and relatively inexpensive, may offer the best alternative source of covariates for SAE in this context. In this paper, we present area-level models involving covariates obtained from spatial big data

sources, such as remotely-sensed data (e.g., night time lights, land cover, elevation), climatic data (e.g., temperature, precipitation) and GIS data (e.g., road networks), with the aim to explore the opportunities and methodological challenges associated with their use in SAE.

193 Measuring well-being at local level using big data

Ralf Münnich

Universität Trier, Deutschland
muennich@uni-trier.de

SECTION: MAKSWELL EU

Health and well-being become more and more important within the framework of sustainable development goals (SDG) as a measure for the population. The current SDG indicators measure the different goals by global measures on country-level. In many cases, local structures and local differences play an important role, especially in the context of the SDG 3, good health and well-being. SDG 3 includes 13 targets and in total 26 proposed indicators to measure and monitor good health and well-being. As explained in World Health Organization (2016) a crucial challenge in collecting information with these indicators is not only the calculation of global indicators but also coverage indicators which shape a more precise picture on the local risks and obstacles towards this sustainable development goal. The success in reaching for example target 3.9 „substantially reduce the number of deaths and illnesses from hazardous chemicals and air, water and soil pollution and contamination“ is likely to depend on the identification of critical areas and the interaction between pollution and health as well as other influencing factors for the population living. People living in industrial areas or residential areas close to polluted water sources might show highly different results regarding the indicators as the overall country measures. Disaggregated results for the indicators will therefore help to identify disadvantaged population within a country (World Health Organization 2016, p. 4 and p. 29). In this paper, different approaches to measure selected well-being indicators on regional and local level are discussed and further developed. Infrastructural information from geographic data sources is employed to elaborate local patterns of indicators for wellbeing. These data comprise flight and satellite data. Census data will be used, provide information on grid-level. Finally, survey data will be integrated, which, in general, are available only on administrative levels, and hence, cover larger areas especially in town areas.

194 Welche Chancen und Hürden bietet die Digitalisierung im schulischen Bildungskontext? – Ein Datenjournalismusprojekt

Susanne Kirschstein-Barczewski

Martin-Luther Universität Halle-Wittenberg, Deutschland
s.kirschstein.barczewski@gmail.com

SECTION: Data Journalism

Schwache Ergebnisse in internationalen Vergleichen (PISA) und steigender Fachkräftemangel aufgrund mangelnder Qualifizierung von Absolventen der Sekundarstufe 1, sich verändernde Anforderungen an die Absolventen und neue politische Ansprüche stellen das System immer wieder vor neue Herausforderungen. Eine dieser Herausforderungen ist die durch die Bundesregierung im DigitalPakt Schule beschlossene Digitalisierung der Schulen.

Obwohl die Digitalisierung der Schulen bisher kaum angelaufen ist, gibt es erste Hinweise darauf, dass der Gebrauch moderner Medien einen negativen Einfluss auf die Fähigkeiten und Fertigkeiten der deutschen Schüler hat. Immer wieder stehen die abnehmende Konzentrations- und Merkfähigkeit sowie mangelnde Kreativität der Schüler im Mittelpunkt. Den identifizierten Problemen stehen steigende Zahlen an Abiturienten sowie steigende Zahlen an Studierenden und jungen Menschen mit einem Hochschulabschluss gegenüber.

Verschiedene Studien haben punktuell gezeigt, welche Einflüsse auf das erfolgreiche Abschließen der Schulischen Bildung es gibt. Berücksichtigung finden oft sozio-ökonomische Aspekte wie der soziale Hintergrund und/oder ein Migrationshintergrund. Die Arbeitsräume der Schüler, Lehrer und anderer beteiligter Erziehungs- und Lehrpersonen, deren Aus- und Weiterbildung sowie die technische Ausstattung, das Konzept und die Form der jeweiligen Schulen finden hingegen kaum oder keine Beachtung. Unterschiedliche Erfolge in der weiteren Ausbildung der Absolventen unterschiedlicher Schulen lassen allerdings einen relevanten Einfluss solcher Einflussfaktoren vermuten.

Gründe für die mangelnde Berücksichtigung von Einflussgrößen können in der Verfügbarkeit von Daten liegen. Sowohl die politische, als auch die individuell empfundene Sensibilität der Thematik, erschweren die Erfassung benötigter Größen. Die Analyse konkreter Bildungsverläufe auf Basis zuordenbarer persönlicher Daten und damit eine gezielte Identifikation möglicherweise problematischer Konstellationen von Einflüssen, erweist sich als schwierig.

Um konkrete Daten- und Informationsbedarfe abzuleiten, werden in einem ersten Schritt relevante Datenquellen recherchiert und eine Bestandsaufnahme verfügbarer Daten vorgenommen. Hieraus können in einem zweiten Schritt Bedarfe an zur Zeit noch nicht verfügbaren Daten abgeleitet und Möglichkeiten zu deren Beschaffung und Bereitstellung diskutiert werden.

195 Daten-Kompetenz für Journalisten – das Dortmunder Modell

Paul Klammer; Holger Wormer

TU Dortmund, Germany
paul.klammer@tu-dortmund.de

SECTION: Data Journalism

Journalisten werden in ihrem Alltag fortlaufend mit Daten konfrontiert. Ob Befragungen, Marktanalysen oder wissenschaftliche Publikationen – ohne ein Grundverständnis von Statistik ist keine unabhängige Berichterstattung möglich. Zugleich werten Datenjournalisten selbstständig Datensätze aus oder erheben sie gleich in Eigenregie oder in Zusammenarbeit mit Wissenschaftlern. Diese Entwicklung stellt auch die Journalistenausbildung vor eine Herausforderung. Welche Kompetenzen müssen in den Redaktionen für eine qualitativ hochwertige Berichterstattung vorhanden sein? Was muss jeder Journalist mindestens wissen? Wo sind Journalisten auf die Unterstützung von Statistikern angewiesen? Der Vortrag stellt ein mehrstufiges Konzept für die Ausbildung von Journalisten vor, wie es am Institut für Journalistik der TU Dortmund angewendet wird – von Grundkenntnissen bis zum eigenständigen Studienschwerpunkt Datenjournalismus.

196 Wozu brauchen Datenjournalisten „Statistical Literacy“ und welchen Mehrwert können interaktive Visualisierungen für Online-Publikationen haben?

Hans-Joachim Mittag (1); Katharina Schüller (2)

1: University of Hagen, Hagen, Germany; 2: Stat-Up, Munich, Germany
joachim.mittag@fernuni-hagen.de

SECTION: Data Journalism

Der Datenjournalismus hat sich als eigenständiger Teilbereich des Online-Journalismus etabliert. Datenjournalisten schreiben erzählungswürdige Geschichten, die sich auf valide Daten stützen. Sie suchen Botschaften, die in großen Datenmengen oft nicht unmittelbar sichtbar sind und nutzen dabei unterschiedliche Werkzeuge. Im Zuge der fortschreitenden Digitalisierung ist das verfügbare Datenangebot kaum noch zu überschauen.

Das Tagesgeschäft von Datenjournalisten erfordert ressortübergreifende Kompetenzen. Es gilt geeignete Daten zu finden, zu validieren, zu analysieren und nach Auffälligkeiten zu durchsuchen, bevor sie etwa in Form von Infografiken oder mit interaktiven Karten veranschaulicht werden. Schließlich ist eine auf den Daten aufbauende Geschichte zu entwickeln und termingerecht zu publizieren. Im Vortrag wird diskutiert, wie wichtig bei all diesen Arbeitsschritten insbesondere „Statistical Literacy“ ist.

Im Datenjournalismus findet man häufiger Beiträge mit interaktiven Visualisierungen. Deren potenzieller Mehrwert für die Leser wird aber meist kaum ausgeschöpft. So können Leser selten verschiedene Sichten auf einen vorgegebenen Datensatz realisieren und dabei unterschiedliche grafische Instrumente einsetzen. In einem mit dem Reporterpreis 2017 ausgezeichneten datenjournalistischen Beitrag von „Zeit

Online“ bestand die einzige Interaktivität z.B. darin, dass Lesern einfache Ja-Nein-Fragen präsentiert wurden und sie zu ihrer Auswahl eine Wahr-Falsch-Rückmeldung mit einem Säulendiagramm erhielten.

Dieses Defizit liegt darin begründet, dass Online-Redaktionen alle Komponenten einer Publikation bevorzugt hausintern und möglichst weitgehend automatisiert erstellen. Präferiert werden Open-Source-Werkzeuge, die keine Programmierkenntnisse erfordern. Maßgeschneiderte Programmierung gilt als zu teuer und zu aufwändig. Die Argumente sind nachvollziehbar, greifen aber nicht bei Datensätzen, die in festen Abständen aktualisiert werden, ohne ihren Aufbau zu verändern. Hier könnte mit deutlich anspruchsvolleren Templates gearbeitet werden, in die regelmäßig neue Daten eingespeist werden. Das Statistische Bundesamt liefert hierzu gutes Anschauungsmaterial, z.B. animierte Darstellungen der Bevölkerungsentwicklung und des Verbraucherpreisindex.

Im Vortrag werden interaktive Visualisierungen präsentiert, die eine ansatzlose Exploration von Datensätzen mit unterschiedlichen grafischen Instrumenten bieten. Dabei werden Daten des Kraftfahrtbundesamts, des Friedensforschungsinstituts SIPRI und Eurostat verwendet.

197 Bayesian and Spline based Approaches for (EM based) Graphon Estimation

Göran Kauermann; Benjamin Sischka

Ludwig-Maximilians-Universität München, Germany
benjamin.sischka@stat.uni-muenchen.de

SECTION: Bayesian Analysis in Natural and Social Science

We propose a new method for graphon estimation using the principles of the EM algorithm. The approach considers both, variability with respect to ordering the nodes of a network and estimation of the unique representation of a graphon. To do so (linear) B-splines are used, which allows to easily accommodate constraints in the estimation routine. Those are necessary to ensure that the estimated graphon possesses essential features and fulfills the canonical representation, meaning its univariate margin is monotonic. The graphon estimate itself allows to apply Bayesian ideas to achieve posterior information and to explore the plausibility of reordering the nodes. Variability and uncertainty is taken into account using MCMC techniques. Combining both steps leads to an EM based approach for graphon estimation, which yields under certain circumstances much better results than previous methods.

198 Non-Informative Bayesian Inference for Heterogeneity in a Generalized Marginal Random Effects Meta-Analysis

Olha Bodnar

Örebro University, Deutschland
olha.bodnar.boretska@gmail.com

SECTION: Bayesian Analysis in Natural and Social Science

This presentation describes an objective Bayesian inference procedure for the heterogeneity parameter in a generalized marginal random effects model. Models of this kind are widely used in meta-analysis and in inter-laboratory comparisons. Assuming that the data can be adequately modeled using elliptically contoured distributions, we derive a reference prior for the model parameters, and produce an analytical expression of the corresponding posterior. We also state necessary conditions for the resulting posterior to be proper and for its first two moments to be finite. The general theoretical results are illustrated for three well-known families of elliptically contoured distributions: normal, Student's t , and Laplace.

199 Bayesian Integration of Probability and Non-Probability Samples: Implications for Survey Estimation and Survey Costs

Joseph Sakshaug (1,3); Arkadiusz Wisniowski (2); Diego Perez-Ruiz (2); Annelies Blom (3)

1: Institute for Employment Research (IAB); 2: University of Manchester; 3: University of Mannheim
joe.sakshaug@iab.de

SECTION: Bayesian Analysis in Natural and Social Science

Web surveys conducted via volunteer panels have become a dominant method of collecting survey data. As no interviewers are involved, these panels offer a distinctively cheaper and faster means of data collection compared to more traditional modes. However, because volunteer online panels do not utilize probability sampling, concerns prevail that such panels are inferior to their probability-based counterparts in terms of meeting the high quality standards expected of scientific research. Indeed several comparative studies have shown that samples drawn from volunteer panels are less representative than samples drawn using probability-based methods. Given that the popularity of volunteer samples is unlikely to abate, it is prudent to consider ways of integrating them with probability samples in order to overcome their weaknesses. We contribute to this notion by implementing a Bayesian approach that integrates both sampling types. We demonstrate the method using nationally-representative probability-based samples collected in Germany alongside eight concurrently collected non-probability samples from different survey vendors. To evaluate the method, the error properties of several survey estimates derived from the integrated samples are compared with estimates derived from non-integrated probability-only samples. We show that the integrated approach yields estimates with smaller mean-squared error (MSE) compared to the standard (non-integrated) approach. Moreover, based on a cost analysis using actual cost data from the non-probability samples, we show that integrating both probability and non-probability samples can potentially yield substantial cost savings relative to utilizing probability-only samples.

200 Rate-efficient estimation of the BG index of a Lévy process

Fabian Mies

RWTH Aachen University, Germany
mies@stochastik.rwth-aachen.de

SECTION: Statistics in Finance

The Blumenthal–Gettoor (BG) index characterizes the jump measure of an infinitely active Lévy process. It determines sample path properties, and affects the behavior of many econometric procedures. Based on high-frequency observations, estimation of the BG index is in general possible only at a rate slower than the parametric square-root, unless the model is a pure-jump process. However, existing estimators only achieve rates of convergence which are suboptimal by a small polynomial factor.

In this paper, a novel estimator for the BG index is presented, which is rate-optimal up to logarithmic factors. Furthermore, our method yields a new volatility estimator which accounts for jumps of infinite variation. A simulation study compares the finite sample behavior of the proposed estimators to competing methods from the financial econometrics literature.

201 Encompassing Tests for Higher-Order Elicitable Functionals

Timo Dimitriadis (2,3); Julie Schnaitmann (1)

1: Universität Konstanz, Germany; 2: Universität Hohenheim; 3: Heidelberg Institute for Theoretical Studies HITS gGmbH
julie.schnaitmann@uni-konstanz.de

SECTION: Statistics in Finance

We introduce encompassing tests for forecasts of higher-order elicitable functionals such as the variance and the Expected Shortfall (ES). Classical encompassing tests rely on the existence of strictly consistent loss functions for the forecasted functionals under consideration. For the aforementioned functionals, such strictly consistent loss functions only exist for the pairs (mean, variance) and (quantile, ES). We focus on proposing encompassing tests for the ES, where we propose three test variants. The first tests joint encompassing of the ES together with the quantile, whereas the second and third test encompassing of the ES stand-alone. These latter two are different as the second one needs auxiliary quantile forecasts whereas the third variant works without them and can be applied in situations where only ES forecasts are available. These tests provide a theoretical justification for forecast combination of the ES when encompassing is rejected. We show through simulation studies that all proposed tests are reasonably sized and exhibit good power properties against general alternatives in typical financial applications. In the empirical application, we apply encompassing tests in order to demonstrate the superiority of forecast combination methods for the ES.

202 Stochastic Volatility Jump Diffusion Models: Moments, Estimation and Simulation

Manuel Schmid (1); Ostap Okhrin (1); Michael Rockinger (2)

1: TU Dresden, Germany; 2: University of Lausanne, Switzerland
manuel.schmid@tu-dresden.de

SECTION: Statistics in Finance

In this paper we consider a novel stochastic volatility jump diffusion model for log-returns. We determine explicit formulas for the unconditional moments, up to the fourth order, and for temporal measures of dependence of the stochastic variance process and the log-price return process. We show how to estimate the model via MCMC and EMM. We develop a sampling scheme capable of generating observations which match the first three moments of the model. In an empirical section we reconsider estimators for various moments and determine the convergence properties for given sample sizes.

203 An approximate solution for the power utility optimization under predictable returns

Taras Bodnar (2); Dmytro Ivasiuk (1); Nestor Parolya (3); Wolfgang Schmid (1)

1: European University Viadrina; 2: Stockholm University; 3: Delft University of Technology
ivasiuk@europa-uni.de

SECTION: Statistics in Finance

In this work we present an approximate solution of the portfolio choice problem for an investor with a power utility function. Assuming that the asset returns follow a vector autoregressive process with normally distributed error terms it follows that portfolio gross returns become normally distributed. We show that, the log-normal distribution seems to be a good proxy of the normal distribution if the standard deviation of the normal distribution is much smaller than its mean. Such a situation can be frequently observed in finance and this fact is used to derive the optimal weights. In addition, we provide a simulation study comparing the derived result to the well-known numerical solution obtained by using a Taylor series expansion of the value function.

204 Using Bayesian Additive Regression Trees for Modelling Attrition and Unit Non-Response in Longitudinal Surveys

Timo Gnams (2,3); Sabine Zinn (1,2)

1: SOEP DIW, Berlin, Germany; 2: Leibniz Institute for Educational Trajectories, Bamberg, Germany; 3: Johannes Kepler University, Linz, Austria
szinn@diw.de

SECTION: Methodology of Statistical Surveys

The analysis of attrition and unit non-response is a standard procedure for the evaluation of the quality of survey data and for the construction of adjustment factors for survey weights to compensate for the selective loss of cases. To examine the process of non-response and attrition, typically, logit or probit models are used, either in their plain versions if the model outcome is purely binary (i.e., participation yes or no) or as multinomial models if the outcome is categorical (e.g., participation, temporary dropout, final dropout). Commonly all conceivable variables (including their interactions) that may impact the participation propensity are used as explanatory variables in the respective logit or probit model. In large-scale surveys such as the Socio-Economic Panel Study (SOEP) or the National Educational Panel Study (NEPS) it is common practice to study the impact of a plethora of variables on the non-response behaviour of the targets. This strategy often yields oversized regression models with weakly occupied cells. To overcome this problem, we suggest using non-standard regression trees instead: Bayesian Additive Regression Trees (BART). BART is a flexible sum of trees approach that overcomes problems of standard regression trees such as prior specification issues or the independence assumptions of random forests. It implements an iterative Bayesian backfitting MCMC algorithm that generates samples from a posterior. We use BART to study the participation of students in the longitudinal NEPS across five waves. For this purpose, we model first the students' propensity to remain in one of the schools surveyed by the NEPS using a survival model. Thereafter, we study their probability for recurrent participation in the panel waves of the NEPS by means of an event history approach. We conclude our talk by contrasting the results that we got by the BART models with the results of a classical logit approach.

205 Longitudinal weights and attrition in the EU-Labour Force Survey

Hannah Kiiver; Frank Espelage

Eurostat, Luxembourg
hannah.kiiver@ec.europa.eu

SECTION: Methodology of Statistical Surveys

The EU-Labour Force Survey (EU-LFS) is designed as a quarterly cross-sectional survey. Among the most important indicators published based on the collected data are quarterly and annual estimates of the three labour market statuses unemployment, employment and inactivity. Since almost all countries use rotational patterns with quarterly and annual sample overlap, Eurostat has started to derive estimates for

the 9 labour market status transitions based on these overlapping sub-samples. Longitudinal weights are derived for this sub-sample such that transition estimates and indicators based on the full cross-sectional samples are consistent. First, weights are scaled to meet the target period margins for ILO labour market status by sex and 10-year age group. In a second step, transition matrices between the labour market statuses in the initial and target period for the complete 15-74 year age group are aggregated based on the reweighted data; final longitudinal weights are consequently derived from an iterative raking procedure .

This paper investigates whether estimates based on these longitudinal weights for annual transitions might be biased by differential attrition, and if adding further variables in the reweighting procedure before the raking step changes the results significantly. The analysis investigates the factors associated with increased probability for sample drop out for each country individually. Personal characteristics are in most countries strong determinants of the probability to remain in the sample. Some of the variables identified as relevant for individual attrition probability are included in the reweighting step, and longitudinal indicators of both approaches are compared. The discussion of these results show that a large increase in the number of variables used in reweighting entails some risk, and that using the same approach for all countries delivering EU-LFS data may actually lead to increases in existing bias in some cases.

206 From Field Survey to Land Use Statistics – Supporting the LUCAS project

Christian Joebges

GOPA Luxembourg, Luxembourg
christian.joebges@gopa.lu

SECTION: Methodology of Statistical Surveys

The Land Use/Cover Area frame Survey (LUCAS) carried out by Eurostat provides harmonized data for the EU-28 member states every 3 years since 2009. LUCAS data is used for different official statistics (e.g. SDGs) and therefore needs to provide versatile data. Spread over 337.854 points across Europe, up to 70 values are recorded for each point location of the sample to indicate land cover, land use and corresponding parameters for land management. Since the initial survey in 2009 the amount of data, the collection of variables and the overall complexity of LUCAS were constantly growing. To handle the immense workload Eurostat was supported by GOPA Consultants since 2014 for stratification of the sample, validation of the data and calculation of estimates. Since January 2019, right after the data collection for LUCAS 2018 was almost finished; a new project of 3 years duration started which represents a more wide scale approach to achieve much needed improvements. The methodological support offered by GOPA Luxembourg in the frame of this contract targets the coherence and comparability of the micro data across the different survey rounds, the quality of the procedures used for the data collection and the correlation of LUCAS data with other national and EU sources. For this purpose the data collection gets evaluated, validated, cleaned and imputed statistically. In addition, recommendations will be provided for increasing the value of the actual and future LUCAS results towards the generic public use and other EU institutions or agencies. Results and conclusions drawn from all the 11 Tasks get finally completed with a SWOT analysis and the proposal of potential enhancements for future surveys. The extent of this methodological assistance to Eurostat and the approach of combining statistical knowledge with ecological expertise mark a milestone for the development and improvement of LUCAS survey.

207 How to Find Him or Her Again?

Sigbert Klinke (1); Jasmin Schlax (2)

1: Humboldt-Universität zu Berlin, Germany; 2: Johannes-Gutenberg-Universität Mainz, Germany
sigbert@wiwi.hu-berlin.de

SECTION: Methodology of Statistical Surveys

In the WiWiKom II Project (Entwicklung der wirtschaftswissenschaftlichen bzw. gesellschaftsökonomischen Fachkompetenz im Verlauf des Studiums), a longitudinal study is conducted with students using one questionnaire per term (semester). Due to Germany's strict privacy policies and Europe's General Data Protection Regulation we were not permitted to simply collect student identities to find the same student in different waves.

Instead of student IDs we use a nearly unambiguously alphanumeric code, and the email address (provided that the student agrees to it) for re-identification. The alphanumeric code is also used to give the students feedback about their results in comparison with reference groups (male/female, subject of study).

Using the Levenshtein distances between codes and emails we generate possible observation matches between waves. Besides the code or the email, other variables (age, sex, etc.) can be used to pre-evaluate a possible match. The possible matches are exported to MS-Excel files and can be evaluated, for example by the inspection of the questionnaires. We have to balance the number of inspections with the number of possible matches. A matched data set is built from the corrected matched files.

For the WiWiKom II project we have developed our R package „findMatch“; (<https://github.com/sigbertklinke/findMatch>) which contains functions to help find observations in two waves (findMatch). These two wave matches are then used to find observations in three or more further waves (findPanel).

208 Introduction to the project MAKSWELL

Fabio Bacchini; Maria Grazia Calza

Istat, Italy
bacchini@istat.it; calza@istat.it

SECTION: MAKSWELL EU

The project MAKSWELL (MAKING Sustainable development and WELL-being frameworks work for policy) proposes to extend and harmonising the indicators able to capture the main characteristics of the beyond-GDP approach proposing a new framework that includes them in the evaluation of the public policies. The main goals of the project can be summarized in three main objectives: building up a database for a wide set of EU countries that select and harmonize the national framework on wellbeing as well as the available SDG indicators; improving the Database both in relation to the timeliness and to the integration with big data measures and the methodologies able to reach these extensions; using the extended Database for policy evaluation and built up a national pilot studies.

The analysis of the country reports showed that 19 countries over 28 developed a framework on well-being, in 11 cases they are used in the national policy cycle. For 12 countries indicators are available below the national level. As far as the SDGs are concerned, the analysis of the country reports shows that all countries implemented, or are about to implement (Spain), the Agenda 2030. In 21 cases SDGS indicators are used in the national policy cycle. For 12 countries SDGs indicators are available below the national level.

Traditionally sample surveys are the data source used for measurement frameworks for well-being and sustainability. However over the last decades alternative data sources become available such as scanner data, mobile phone, web scraping, social media, satellite data. A review of this new data sources and their potentiality for the project MAKSWELL have been explored in the deliverable 2.1 while deliverable 2.2 have illustrated the combination amid survey data and non-traditional data sources.

Amid the indicators of well-being those ones related to poverty are extremely important. WP3 aims to explore the methodology useful to improve the availability of these indicators at regional level.

Interaction of new data and methodologies to improve timeliness and time series of the indicators are explored in WP4 while WP5 This work package has the objective to test a single Beyond-GDP framework effective for guiding policy making at local and national level integrating micro and macro statistical information. The activities on these 2 WP are currently ongoing.

Participants – Teilnehmer:

Ahlborn, Markus; Statistisches Bundesamt; markus.ahlborn@destatis.de
Alhorn, Kira; TU Dortmund; kira.alhorn@tu-dortmund.de
Andrä, Diana; Stadt Dortmund/Dortmunder Statistik; dandrae@stadtdo.de
Articus, Charlotte; Trier University; articus@uni-trier.de
Bachmann, Günther; Amt für Wirtschaft und Stadtentwicklung; guenther.bachmann@darmstadt.de
Bahromov, Jamol; Ruhr Universität Bochum; jamol.bahromov@rub.de
Ballreich, Dominik; codecentric AG; dominik.ballreich@gmail.com
Barunik, Jozef; Charles University in Prague; Institute of Economic Studies; barunik@fsv.cuni.cz
Baumgärtner, Luisa; Statistisches Bundesamt; luisa.baumgaertner@destatis.de
Beck, Martin; Statistisches Bundesamt; martin.beck@destatis.de
Becker, Janis; Leibniz Universität Hannover; becker@statistik.uni-hannover.de
Becker, Angelika; Verband der Chemischen Industrie e.V.; becker@vci.de
Behmann, Mathias; Stadtmagistrat Innsbruck; mathias.behmann@magibk.at
Bekalarczyk, Dawid; University of Duisburg-Essen; dawid.bekalarczyk@uni-due.de
Berens, Florian; University of Goettingen; florian.berens@uni-goettingen.de
Berger, Theo; University of Bremen; thberger@uni-bremen.de
Biecek, Przemyslaw; Warsaw University of Technology; przemyslaw.biecek@gmail.com
Billé, Anna Gloria; Free University of Bolzano-Bozen; annagloria.bille@unibz.it
Birkigt, Holger; Statistisches Bundesamt; holger.birkigt@destatis.de
Birner, Nadine; VDI/VDE Innovation + Technik GmbH; Nadine.birner@vdivde-it.de
Blätgen, Nadine; BBSR; nadine.blaetgen@bbr.bund.de
Bleninger, Sara; Bayerisches Landesamt für Statistik; sara.bleninger@statistik.bayern.de
Bodnar, Taras; Stockholm University; taras.bodnar@math.su.se
Bodnar, Olha; Örebro University; olha.bodnar.boretska@gmail.com
Bömermann, Hartmut; Amt für Statistik Berlin Brandenburg; diana.freitag@statistik-bbb.de
Bonofiglio, Federico; Albert-Ludwigs-Universität Freiburg; bono@imbi.uni-freiburg.de
Borsi, Lisa; Universität Trier; borsi@uni-trier.de
Bours, Monika; RWTH Aachen University; bours@stochastik.rwth-aachen.de
Braakmann, Albert; Statistisches Bundesamt; albert.braakmann@destatis.de
Breidenbach, Philipp; RWI - Leibniz-Institut für Wirtschaftsforschung; breidenbach@rwi-essen.de
Brenzel, Hanna; Federal Statistical Office (Destatis); Hanna.Brenzel@destatis.de
Breuer, Hermann; Stadt Köln; hermann.breuer@stadt-koeln.de
Brockner, Sven; University Duisburg-Essen; sven.brockner@uni-due.de
Bruckmeier, Kerstin; Institut für Arbeitsmarkt- und Berufsforschung (IAB); Kerstin.Bruckmeier@iab.de
Brügge, Nils; RWTH Aachen University; bruegge@stochastik.rwth-aachen.de
Bücher, Axel; Heinrich-Heine-Universität Düsseldorf; axel.buecher@hhu.de
Buchholz, Christine; Hochschule Bonn-Rhein-Sieg; christine.Buchholz@h-brs.de
Burgard, Jan Pablo; Universität Trier; burgardj@uni-trier.de
Burger, David; Hansestadt Lübeck; david.burger@luebeck.de
Buschner, Andrea; Bayerisches Landesamt für Statistik; andrea.buschner@statistik.bayern.de
Büttner, Tobias; Bundesagentur für Arbeit; tobias.buettner@arbeitsagentur.de
Caratiola, Christopher; University Trier; caratiola@uni-trier.de
Chen, Gong; Technische Universität Dresden; gong.chen1@tu-dresden.de

Chlumsky, Jürgen; Statistisches Bundesamt; juergen.chlumsky@destatis.de
Cohrs, Renate; Statistisches Amt für Hamburg und Schleswig-Holstein; Vorstand@statistik-nord.de
Czarnitzki, Dirk; KU Leuven; dirk.czarnitzki@kuleuven.be
Dashuk, Valerii; Ludwig-Maximilians-Universität (LMU) München; Valerii.Dashuk@econ.lmu.de
Daube, Matthias; Thüringer Landesamt für Statistik; matthias.daube@statistik.thueringen.de
Debald, Yvonne; Stadt Aachen; yvonne.debald@mail.aachen.de
Dechent, Jens; Statistisches Bundesamt; jens.dechent@destatis.de
Dekkers, Gijs; Federal Planning Bureau; gd@plan.be
Delfs, Silke; Statistik der BA; Silke.Delfs@arbeitsagentur.de
Demetrescu, Matei; CAU Kiel; mdeme@stat-econ.uni-kiel.de
Dickopf, Xaver Peter Alfons; Statistisches Bundesamt; xaver.dickopf@destatis.de
Dieckmann, Hanna; Trier University; dieckmann@uni-trier.de
Dimitriadis, Timo; University of Hohenheim; timo.dimi@googlemail.com
Dmytriv, Solomiia; European University Viadrina; dmytriv@europa-uni.de
Dohnke, Jan; Wissenschaftsstadt Darmstadt; jan.dohnke@darmstadt.de
Dorner, Matthias; Institut für Arbeitsmarkt- und Berufsforschung (IAB); matthias.dorner@iab.de
Dörr, Patricia; Universität Trier; doerr@uni-trier.de
Dräger, Sebastian; Trier University; draegers@uni-trier.de
Dumpert, Florian; Statistisches Bundesamt; florian.dumpert@destatis.de
Dürre, Alexander; Technische Universität Dortmund; alexander.duerre@udo.edu
Dyckerhoff, Rainer; Universität zu Köln; rainer.dyckerhoff@statistik.uni-koeln.de
Ebigt, Sascha; Landesamt für Statistik Niedersachsen; Sascha.Ebigt@statistik.niedersachsen.de
Eckendorff, Roman; Statistik der Bundesagentur für Arbeit; roman.eckendorff2@arbeitsagentur.de
Eckert, Olaf; Statistisches Bundesamt; olaf.eckert@destatis.de
Eckl, Verena; SV Wissenschaftsstatistik GmbH; verena.eckl@stifterverband.de
Eichhorn, Lothar; Wissenschaftliche Gesellschaft zum Studium Niedersachsens; lothar.eichhorn@htp-tel.de
Eichhorn, Tanja; LSN; tanja.eichhorn@statistik.niedersachsen.de
Emmenegger, Jana; Destatis; jana.emmenegger@destatis.de
Ertz, Florian; Universität Trier; ertz@uni-trier.de
Espelage, Frank; European Commission; frank.espelage@ec.europa.eu
Espenkotte, Thorsten; Statistik der Bundesagentur für Arbeit; thorsten.espenkotte@arbeitsagentur.de
Feske, Carina; Regionalverband Ruhr; feske@rvr.ruhr
Feuerhake, Joerg; Statistisches Bundesamt; joerg.feuerhake@destatis.de
Filzmoser, Peter; Vienna University of Technology; P.Filzmoser@tuwien.ac.at
Fink, Paul; Bayerisches Landesamt für Statistik; paul.fink@statistik.bayern.de
Fitzenberger, Bernd; HU Berlin; fitzenbb@hu-berlin.de
Fleischer, Karlheinz; Philipps-Universität Marburg; k.fleischer@wiwi.uni-marburg.de
Flores, Luis Federico; Destatis; lfedeflores88@gmail.com
Frahm, Gabriel; Helmut Schmidt Universität; frahm@hsu-hh.de
Franke, Sebastian; Ludwig-Maximilians-Universität München; sebastian.franke@bwl.lmu.de
Fried, Roland; TU Dortmund; fried@statistik.tu-dortmund.de
Frietsch, Rainer; Fraunhofer ISI; rainer.frietsch@isi.fraunhofer.de
Fritsch, Markus; University of Passau; markus.fritsch@uni-passau.de
Friz, Katharina; Universität Bremen; frizkat@uni-bremen.de
Frohn, Christoph; Universität Duisburg-Essen; christoph.frohn@uni-due.de

Fugger, Gerald; Universität-Duisburg Essen; gerald.fugger@uni-due.de
Fürnrohr, Michael; Bayerisches Landesamt für Statistik; michael.fuernrohr@statistik.bayern.de
Garthoff, Robert; Statistisches Landesamt des Freistaates Sachsen; Robert.Garthoff@statistik.sachsen.de
Gaykalov, Konstantin; Europa-Universität Viadrina; gaykalov@europa-uni.de
Germakowsky, Gerd; Stadt Witten; gerd.germakowsky@stadt-witten.de
Gheno, Gloria; Innovative data analysis; gloriagheno@libero.it
Giese, Marco; University of Duisburg-Essen; Marco.Giese@wiwinf.uni-due.de
Glauber, Sebastian; Bayerisches Landesamt für Statistik; sebastian.glauber@statistik.bayern.de
Gleich, Andreas; Amt für Statistik und Stadtforschung; Andreas.Gleich@augzburg.de
Glitza, Beate; Statistisches Bundesamt; beate.glitza@destatis.de
Göddecke-Stellmann, Jürgen; Bundesinstitut für Bau-, Stadt- und Raumforschung (BBSR); juergen.goeddecke@bbr.bund.de
Goebel, Jan; DIW Berlin; jgoebel@diw.de
Gondrom, Sabine; Hochschule Ludwigshafen; sabine.go@gmx.de
Görge, Konstantin; Karlsruher Institut für Technologie; Konstantin.goergen@kit.edu
Gornig, Martin; DIW Berlin; mgornig@diw.de
Goschenhofer, Jann; Fraunhofer IIS; jann.goschenhofer@stat.uni-muenchen.de
Gramlich, Tobias; Statistik Hessen (Hessisches Statistisches Landesamt); tobias.gramlich@statistik.hessen.de
Gribisch, Bastian; Universität zu Köln; bastian.gribisch@statistik.uni-koeln.de
Gronbach, Markus; Stadt Pforzheim; Personal- und Organisationsamt; markus.gronbach@pforzheim.de
Große, Franziska; Landesamt für Statistik Niedersachsen; franziska.grosse@statistik.niedersachsen.de
Güdemann, Laura Maria; University Trier; guedemann@uni-trier.de
Gutfleisch, Ralf; Stadt Frankfurt a.M.; ralf.gutfleisch@stadt-frankfurt.de
Haas, Markus; Universität Kiel; markus.haas@qber.uni-kiel.de
Hadam, Sandra; Federal Statistical Office of Germany; sandra.hadam@destatis.de
Haffner, Marietta; Delft University of Technology; m.e.a.haffner@tudelft.nl
Hagn, Harald; Thüringer Landesamt für Statistik; Harald.Hagn@Statistik.Thueringen.de
Hahn, Rainer; Statistik der BA; Rainer.Hahn@arbeitsagentur.de
Hanck, Christoph; Universität Duisburg-Essen; christoph.hanck@vwl.uni-due.de
Hannappel, Marc; Universität Koblenz-Landau; MarcHannappel@uni-koblenz.de
Harms, Torsten; DHBW Karlsruhe; harms@dhbw-karlsruhe.de
Hartl, Tobias; University of Regensburg; tobias1.hartl@ur.de
Hartmann, Michael; Statistik der Bundesagentur für Arbeit; Michael.Hartmann@arbeitsagentur.de
Hauf, Stefan D.; Statistisches Bundesamt; stefan.hauf@destatis.de
Haug, Thorsten; Statistisches Bundesamt; thorsten.haug@destatis.de
Hausmann, Patrick; Covimo; patrick.hausmann@covimo.de
Haußmann, Michael; Landeshauptstadt Stuttgart; Statistisches Amt; michael.hausmann@stuttgart.de
Haustein, Thomas; Statistisches Bundesamt; thomas.haustein@destatis.de
Hecheltjen, Peter; Universität Trier; hecheltjen@uni-trier.de
Heck, Lukas; Landratsamt Karlsruhe; lukas.heck@landkreis-karlsruhe.de
Heckel, Christiane; BIK ASCHPURWIS + BEHRENS GmbH; heckel@bik-gmbh.de
Hees, Katharina; TU Dortmund; hees@statistik.tu-dortmund.de
Heiden, Anja; Stadt Siegen / Statistikstelle; a.heiden@siegen.de
Heinsohn, Till; Landeshauptstadt Stuttgart; till.heinsohn@stuttgart.de
Held, Tobias; Statistisches Amt der Landeshauptstadt Stuttgart; Tobias.Held@stuttgart.de
Hense, Andreas; University Bonn; ahense@uni-bonn.de

Herrmann, Christoph; Martin-Luther-Universität Halle-Wittenberg; christoph.herrmann@wiwi.uni-halle.de
Hin, Monika; Statistisches Landesamt Baden-Württemberg; monika.hin@stala.bwl.de
Hochgürtel, Tim; Statistisches Bundesamt; tim.hochguertel@destatis.de
Höcker, Philip; Universität Duisburg-Essen; philip.hoecker@uni-due.de
Hoffmann, Elke; Deutsche Zentrum für Altersfragen; elke.hoffmann@dza.de
Hoffmann, Jens; Statistik der Bundesagentur für Arbeit; jens.hoffmann@arbeitsagentur.de
Hoffmann, Hanna; Information und Technik Nordrhein-Westfalen (IT.NRW); hanna.hoffmann@it.nrw.de
Hoffmann, Hermann; Beratung Media und Stichprobe; hermann.hoffmann@hh-sample.de
Horneffer, Birgit; Statistisches Bundesamt; birgit.horneffer@destatis.de
Höse, Steffi; Hochschule Zittau/Görlitz; s.hoese@hszg.de
Hötger, Udo; Kreis Lippe; u.hoetger@kreis-lippe.de
Hürter, Marcel; Statistisches Landesamt Rheinland-Pfalz; praesident@statistik.rlp.de
Huschens, Stefan; TU Dresden; stefan.huschens@tu-dresden.de
Ihden, Tanja; Universität Bremen; tanja.ihden@uni-bremen.de
Ivasiuk, Dmytro; European University Viadrina; ivasiuk@europa-uni.de
Jackisch, Annett; Statistikamt Nord; Annett.Jackisch@statistik-nord.de
Jacobi, Lucas; Landeshauptstadt Stuttgart; lucas.jacobi@stuttgart.de
Jäger, Solveigh; BDI; s.jaeger@bdi.eu
Jäger, Armin; Statistik der Bundesagentur für Arbeit; armin.jaeger3@arbeitsagentur.de
Jakubzik, Mirko Alexander; Technische Universität Dortmund; jakubzik@statistik.tu-dortmund.de
Joebges, Christian; GOPA Luxembourg; christian.joebges@gopa.lu
Johannssen, Arne; University of Hamburg; arne.johannssen@uni-hamburg.de
John, Birgit; Statistisches Landesamt Baden-Württemberg; birgit.john@stala.bwl.de
Jung, Robert; Universität Hohenheim; robert.jung@uni-hohenheim.de
Kahl, Katharina; Universität Trier; kahlk@uni-trier.de
Kalinowski, Michael; Bundesinstitut für Berufsbildung; kalinowski@bibb.de
Kauermann, Göran; LMU München; goeran.kauermann@lmu.de
Kaupert, Madeline; Bundesinstitut für Bau-, Stadt- und Raumforschung; madeline.kaupert@bbr.bund.de
Kayser, Susann; Thüringer Landesamt für Statistik; susann.kayser@statistik.thueringen.de
Kellermann, Janosch; Ruhr-Universität Bochum; janosch.kellermann@rub.de
Kibele, Eva; Statistisches Landesamt Bremen; eva.kibele@statistik.bremen.de
Kiiver, Hannah; European Commission; hannah.kiiver@ec.europa.eu
Kirschstein-Barczewski, Susanne; Martin-Luther Universität Halle-Wittenberg; s.kirschstein.barczewski@gmail.com
Kladroba, Andreas; Stifterverband für die Deutsche Wissenschaft; andreas.kladroba@stifterverband.de
Klammer, Paul; TU Dortmund; paul.klammer@tu-dortmund.de
Kleine, Riepke; Bayerisches Landesamt für Statistik; Riepke.Kleine@statistik.bayern.de
Klinke, Sigbert; Humboldt-Universität zu Berlin; sigbert@wiwi.hu-berlin.de
Klotzky, Jens; Statistik der Bundesagentur für Arbeit; Jens.Klotzky@arbeitsagentur.de
Knapp, Guido; TU Dortmund University; guido.knapp@tu-dortmund.de
Kneib, Thomas; Georg-August-Universität Göttingen; tkneib@uni-goettingen.de
Kniebeler, Simone; Statistisches Landesamt Baden-Württemberg; Simone.Kniebeler@stala.bwl.de
Knoth, Sven; Helmut Schmidt University Hamburg; knoth@hsu-hh.de
Koehlmann, Maren; Statistisches Bundesamt / Federal Statistical Office GER; maren.koehlmann@destatis.de
Konrad, Anne; Universität Trier; konrada@uni-trier.de
Kopp, Johannes; Universität Trier; kopp@uni-trier.de

Koppers, Lars; Karlsruher Institut für Technologie; lars.koppers@kit.edu
Korn, Claudia; Statistik der Bundesagentur für Arbeit; claudia.korn@arbeitsagentur.de
Krämer, Walter; TU Dortmund; walterk@statistik.tu-dortmund.de
Kraus, Petra; Hauptverband der Deutschen Bauindustrie e.V.; petra.kraus@bauindustrie.de
Krause, Joscha; Universität Trier; krause@uni-trier.de
Krieger, Sascha; Statistisches Bundesamt; sascha.krieger@destatis.de
Krische, Stephan; GOPA Luxembourg S.à r.l; statistics@gopa.lu
Krümmer, Johanna; Universität Trier; s4jokrue@uni-trier.de
Kubis, Alexander; Institut für Arbeitsmarkt- und Berufsforschung (IAB); alexander.kubis@iab.de
Lange, Tatjana; HS Merseburg; tanja.28.lange@gmail.com
Langrock, Roland; Bielefeld University; roland.langrock@uni-bielefeld.de
Lauerbach, Teresa; Bundesinstitut für Bau-, Stadt- und Raumforschung; teresa.lauerbach@bbr.bund.de
Lebacher, Michael; LMU - Institut für Statistik; michael.lebacher@stat.uni-muenchen.de
Leerhoff, Holger; Amt für Statistik Berlin-Brandenburg; holger@leerhoff.de
Lehmann, Arne; Landesamt für Statistik Niedersachsen; arne.lehmann@statistik.niedersachsen.de
Lenau, Simon; Universität Trier; lenau@uni-trier.de
Lentz, Tobias; Hessisches Statistisches Landesamt; tobias.lentz@statistik.hessen.de
Lenz, Jürgen; Kreis Mettmann; juergen.lenz@kreis-mettmann.de
Lilienthal, Jona; Technische Universität Dortmund; lilienthal@statistik.tu-dortmund.de
Lindner, Ines; Vrije Universiteit Amsterdam; i.d.lindner@vu.nl
Loboda, Dennis; RWTH Aachen; loboda@stochastik.rwth-aachen.de
Lorenz, Kristin; Freie und Hansestadt Hamburg / Behörde für Wirtschaft; Verkehr und Innovation; kristin.lorenz@bwvi.hamburg.de
Loske, Janina; Statistisches Bundesamt; janina.loske@destatis.de
Lübke, Karsten; FOM; karsten.luebke@fom.de
Lukas, David; Martin Luther Universität Halle Wittenberg; david.lukas@wiwi.uni-halle.de
Lux-Henseler, Barbara; Stadt Nürnberg; barbara.lux-henseler@stadt.nuernberg.de
Mädig, Attina; Landeshauptstadt Stuttgart; Attina.Maeding@stuttgart.de
Mai, Christoph-Martin; Statistisches Bundesamt; christoph-martin.mai@destatis.de
Maier, Tobias; Bundesinstitut für Berufsbildung; tobias.maier@bibb.de
Malcherczyk, Dennis; TU Dortmund; dennis.malcherczyk@tu-dortmund.de
Malinovskaya, Anna; Leibniz Universität Hannover; anna.malinovskaya@ikg.uni-hannover.de
Manecke, Julia; Universität Trier; manecke@uni-trier.de
Mannigel, Alice; Statistikamt Nord; alice.mannigel@statistik-nord.de
Mantinger, Mara; Statistisches Landesamt Baden-Württemberg; mara.mantinger@stala.bwl.de
Marbach, Tina; LHS Stuttgart; Statistisches Amt; tina.marbach@stuttgart.de
Mareckova, Jana; University of Sankt Gallen; jana.mareckova@unisg.ch
Martin, Stella; Institut für Ökonometrie und Wirtschaftsstatistik; stella.martin@wiwi.uni-muenster.de
Massmann, Michael; WHU – Otto Beisheim School of Management; michael.massmann@whu.edu
Masuhr, Andreas; WWU Münster; andreas.masuhr@wiwi.uni-muenster.de
Matthews, Sean; Deloitte GmbH; sematthews@deloitte.de
Mause, Nils; RWTH Aachen University; mause@stochastik.rwth-aachen.de
Mehrhoff, Jens; Deutsche Bundesbank; jens.mehrhoff@bundesbank.de
Meinken, Holger; Statistik der Bundesagentur für Arbeit; holger.meinken@arbeitsagentur.de
Merk, Miryam Sarah; European University Viadrina; merk@europa-uni.de
Merkle, Hariolf; Universität Trier; merkle@uni-trier.de

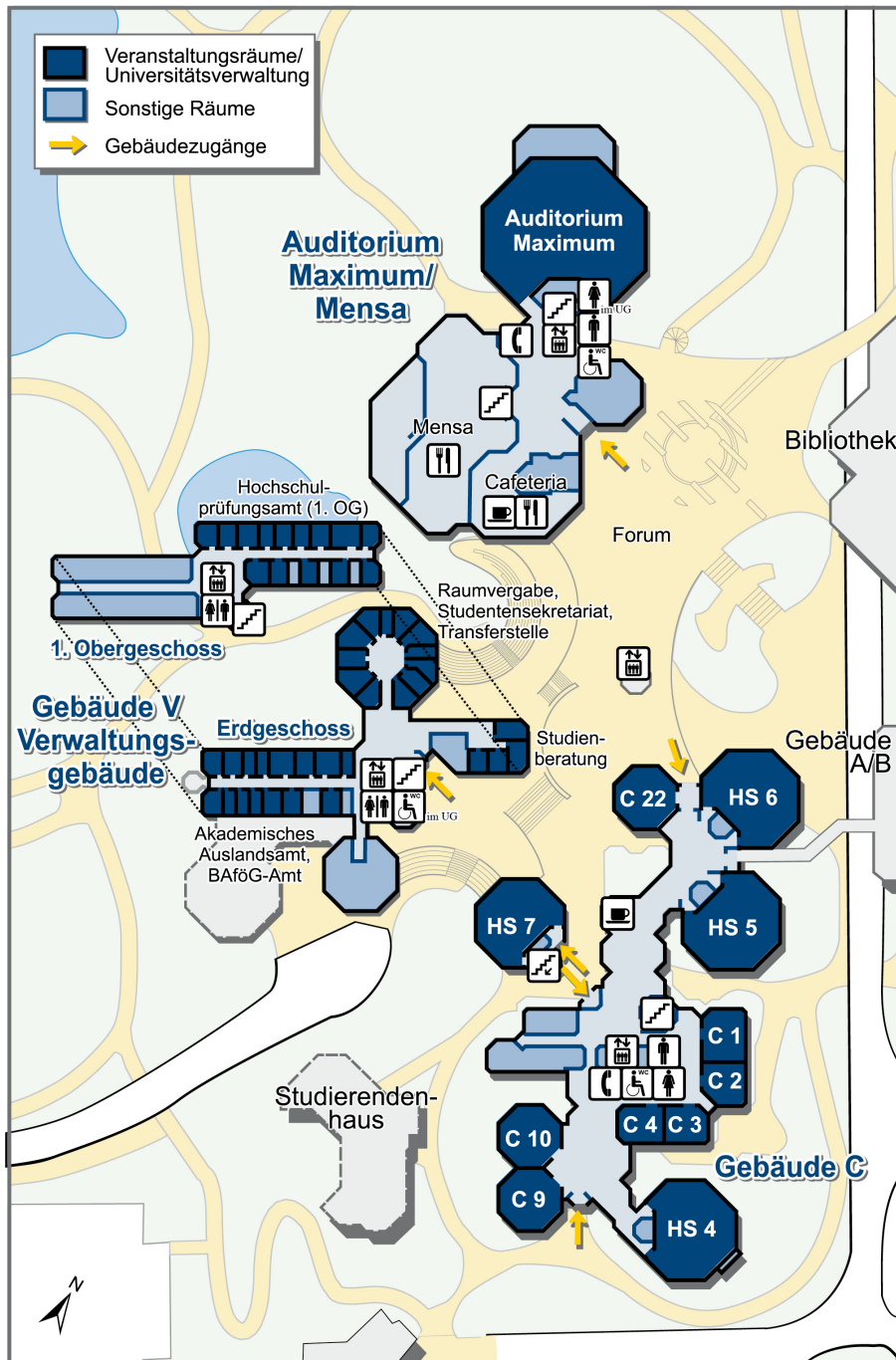
Merz, Joachim; Leuphana Universität Lüneburg; merz@uni.leuphana.de
Mettler, Elena; GOPA Luxembourg; Elena.Mettler@gopa.lu
Meyerhoff, Annette; VDMA e.V.; annette.meyerhoff@vdma.org
Mies, Fabian; RWTH Aachen University; mies@stochastik.rwth-aachen.de
Milbert, Antonia; Bundesinstitut für Bau-, Stadt- und Raumforschung; antonia.milbert@bbr.bund.de
Mittag, Hans-Joachim; University of Hagen; joachim.mittag@fernuni-hagen.de
Morávek, David; University of Economics; davidmoravek@hotmail.com
Mosler, Karl; Universität zu Köln; mosler@statistik.uni-koeln.de
Mozharovskyi, Pavlo; Telecom ParisTech; pavlo.mozharovskyi@telecom-paristech.fr
Mucha, Tanja; Statistisches Bundesamt; tanja.mucha@destatis.de
Mücher, Christian; Universität Konstanz; christian.muecher@uni-konstanz.de
Mühlenfeld, Hans-Ullrich; IT.NRW - Statistisches Landesamt NRW; hans-ullrich.muehlenfeld@it.nrw.de
Mühlenhaupt, Jörg; IT.NRW; Joerg.Muehlenhaupt@it.nrw.de
Mühlhan, Jannek; Institut für Arbeitsmarkt- und Berufsforschung; jannek.muehlhan2@iab.de
Müller, Christine Hildegard; TU Dortmund University; cmueller@statistik.tu-dortmund.de
Münnich, Ralf; Universität Trier; muennich@uni-trier.de
Neuber-Pohl, Caroline; Bundesinstitut für Berufsbildung; neuber-pohl@bibb.de
Neumann, Uwe; RWI - Leibniz-Institut für Wirtschaftsforschung; uwe.neumann@rwi-essen.de
Neurauter, Stefanie; LH Wiesbaden; stefanie.neurauter@wiesbaden.de
Neureiter, Karin; Land Kärnten; neureiter.karin@gmx.at
Nirschl, Thomas; Stadt Nürnberg; thomas.nirschl@stadt.nuernberg.de
Nitt-Drießelmann, Dörte; HWWI Hamburgisches Weltwirtschaftsinstitut; nitt-driesselmann@hwwi.org
Nowossadeck, Enno; Robert Koch-Institut; NowossadeckE@rki.de
Obersneider, Monika; University of Duisburg-Essen; monika.obernsneider@uni-due.de
Ochsmann, Andrea; Stadt Recklinghausen; andrea.ochsmann@recklinghausen.de
Oesting, Marco; University of Siegen; oesting@mathematik.uni-siegen.de
Okasa, Gabriel; University of St.Gallen; gabriel.okasa@unisg.ch
Okhrin, Yarema; Universität Augsburg; yarema.okhrin@wiwi.uni-augsburg.de
Olbricht, Walter; Universität Bayreuth; walter.olbricht@uni-bayreuth.de
Old, Oliver; FernUniversität in Hagen; oliver.old@fernuni-hagen.de
Orlowski, Miriam; Bayerisches Landesamt für Statistik; miriam.orldowski@statistik.bayern.de
Oruc, Hasan Hüseyin; FernUniversität in Hagen; hasan-hueseyin.oruc@fernuni-hagen.de
Ostermann, Fabian; Universität Trier; s4faoste@uni-trier.de
Otto, Anna-Liesa; DStatG; aotto@europa-uni.de
Otto, Philipp; Leibniz University Hannover; otto@ikg.uni-hannover.de
Parolya, Nestor; Delft University of Technology; n.parolya@tudelft.nl
Parton, Frederik; Gesellschaft für wirtschaftliche Strukturforchung mbH; parton@gws-os.com
Pauly, Marcel; Spiegel Online; marcel.pauly@spiegel.de
Peters, Normen; Hessisches Statistisches Landesamt; npeters@statistik.hessen.de
Peters, Bettina; ZEW - Leibniz-Zentrum für Europäische Wirtschaftsforschung; b.peters@zew.de
Petrucci, Alessandra; Università degli Studi di Firenze; alessandra.petrucci@unifi.it
Pflaumer, Peter; TU Dortmund; peter.pflaumer@uni-dortmund.de
Plagens, Manfred; Stadt Würzburg; manfred.plagens@stadt.wuerzburg.de
Pohlmeier, Winfried; Universität Konstanz; Winfried.Pohlmeier@uni-konstanz.de
Porst, Michael; Robert Koch-Institut; porstm@rki.de

Posselt, Mänz; Philipps-Universität Marburg; posselt@wiwi.uni-marburg.de
Pratesi, Monica; University of Pisa; monica.pratesi@unipi.it
Pullen, Markus; LKA Niedersachsen; markus.pullen@polizei.niedersachsen.de
Putscher, Dorothee; Bayerisches Landesamt für Statistik; dorothee.putscher@statistik.bayern.de
Pütz, Peter; Georg-August-Universität Göttingen; ppuetz@uni-goettingen.de
Rabenau, Kai; Fujitsu TDS GmbH; kai.rabenau@tds.fujitsu.com
Radermacher, Walter J.; La Sapienza University Rome; wjr@outlook.de
Rahmenführer, Jörg; Technische Universität Dortmund; rahnenfuehrer@statistik.tu-dortmund.de
Rebien, Martina; Amt für Stadtforschung und Statistik; martina.rebien@web.de
Reeh, Klaus; ex-Eurostat; reeh-oberbillig@t-online.de
Reinhold, Julian; Universität Duisburg-Essen; julian.reinhold@uni-due.de
Rengers, Martina; Statistisches Bundesamt; martina.rengers@destatis.de
Richter, Ann-Kathrin; IT.NRW Statistisches Landesamt; ann-kathrin.richter@it.nrw.de
Richter, Ernst-Joachim; ehemals Stadt Oberhausen; ejochen.richter@arcor.de
Richter, Roland; Stadt Duisburg; R.Richter@Stadt-Duisburg.de
Riede, Thomas; Statistisches Bundesamt; thomas.riede@destatis.de
Rieger, Jonas; Technische Universität Dortmund; jonas.rieger@tu-dortmund.de
Rieke, Jens; Statistisches Landesamt Baden-Württemberg; jensrieko@web.de
Rockmann, Ulrike; Senatsverwaltung für Inneres und Sport Berlin; ulrike.rockmann@uni-oldenburg.de
Rohde, Johannes; IT.NRW; johannes.rohde@it.nrw.de
Rojas-Perilla, Natalia; Freie Universität Berlin; natalia.rojas78@fu-berlin.de
Römer, Ellen; Statistisches Bundesamt; ellen.roemer@destatis.de
Rommelspacher, Simon; Statistisches Bundesamt; simon.rommelspacher@destatis.de
Rosenski, Natalie; Statistisches Bundesamt; natalie.rosenski@destatis.de
Rosteck, Veronika; Springer-Verlag GmbH; veronika.rosteck@springer.com
Rothe, Thomas; Institut für Arbeitsmarkt- und Berufsforschung; thomas.rothe@iab.de
Rudloff, Karin; Statistik der Bundesagentur für Arbeit; karin.rudloff@arbeitsagentur.de
Ruhmann, Iris; Springer-Verlag GmbH Heidelberg; iris.ruhmann@springer.com
Rupp, Martin; Trier University; ruppm@uni-trier.de
Sahir, Ibtissame; GOPA Luxembourg; christiane.gonzalez@gopa.lu
Salish, Nazarii; University Carlos III de Madrid; nsalish@eco.uc3m.es
Salwiczek, Christian; Statistisches Amt für Hamburg und Schleswig-Holstein; christian.salwiczek@statistik-nord.de
Schäfer, Josef; Private; jugschaefer@aol.com
Schäfer, Wolf; Amt für Stadtforschung und Statistik für Nürnberg und Fürth; wolf.schaefer@stadt.nuernberg.de
Schartner, Clara Maria; Destatis; clara.schartner@destatis.de
Schelhase, Torsten; Statistisches Bundesamt; torsten.schelchase@destatis.de
Schels, Helmut; Stadt Ingolstadt; helmut.schels@ingolstadt.de
Schieck, Matthias; Stadt Chemnitz; matthias.schieck@stadt-chemnitz.de
Schierholz, Malte; Bundesagentur für Arbeit; malte.schierholz@iab.de
Schmal, Friederike; Westfälische Wilhelms-Universität Münster; friederike.schmal@wiwi.uni-muenster.de
Schmaus, Simon; Universität Trier; schmaus@uni-trier.de
Schmid, Manuel; TU Dresden; manuel.schmid@tu-dresden.de
Schmid, Wolfgang; Europa-Universität Frankfurt (Oder); schmid@europa-uni.de
Schmidt, Rolf; Rolf Schmidt; schmidt.Ob@gmx.de
Schmidt, Katrin; Statistik der Bundesagentur für Arbeit; katrin.schmidt8@arbeitsagentur.de

Schmidt, Peter; Statistisches Bundesamt; peter.schmidt@destatis.de
Schmitz, Rose; Statistik der Bundesagentur für Arbeit; rose.schmitz@arbeitsagentur.de
Schmitz-Veltin, Ansgar; Landeshauptstadt Stuttgart; ansgar.schmitz-veltin@stuttgart.de
Schneble, Marc; Ludwig-Maximilians-Universität München; marc.schneble@stat.uni-muenchen.de
Schneeweiß, Hans; LMU; schneew@stat.uni-muenchen.de
Schneider, Gaby; Goethe University Frankfurt; schneider@math.uni-frankfurt.de
Schnitzer, Manuel; Statistik der Bundesagentur für Arbeit; Manuel.Schnitzer@arbeitsagentur.de
Schnurbus, Joachim; University of Passau; joachim.schnurbus@uni-passau.de
Schoch, Tobias; University of Applied Sciences Northwestern Switzerland; tobias.schoch@fhnw.ch
Schöneich, Cordula; Statistisches Bundesamt; cordula.schoeneich@destatis.de
Schönfeld-Nastoll, Ulrike; Stadt Oberhausen; ulrike.schoenfeld@oberhausen.de
Schridde, Henning; Statistik der Bundesagentur für Arbeit; henning.schridde@arbeitsagentur.de
Schüler, Yves; Deutsche Bundesbank; yves.schueler@bundesbank.de
Schüller, Katharina; STAT-UP Statistical Consulting & Data Science GmbH; katharina.schueller@stat-up.com
Schultz, Andrea; Stadt Leipzig; andrea.schultz@leipzig.de
Schulz, Julian; Landesamt für Statistik Niedersachsen; julian.schulz@statistik.niedersachsen.de
Schütt, Fabian; Landeshauptstadt Stuttgart; Statistisches Amt; fabian.schuett@stuttgart.de
Schweikert, Karsten; University of Hohenheim; karsten.schweikert@uni-hohenheim.de
Schwengler, Barbara; Institut für Arbeitsmarkt- und Berufsforschung; Barbara.Schwengler@iab.de
Schwickerath, Marco; Statistisches Bundesamt; marco.schwickerath@destatis.de
Seitz, Stefan; Statistisches Landesamt Baden-Württemberg; stefan.seitz@stala.bwl.de
Sell, Jana Franziska; Universität Bremen; jsell@uni-bremen.de
Semeniuk, Ivan; Europa University Viadrina; semeniuk@europa-uni.de
Shkoza, Livia; University of Konstanz; livia.shkoza@uni-konstanz.de
Singer, Hermann; FernUniversität in Hagen; hermann.singer@fernuni-hagen.de
Singer, Kirsten; Statistik der Bundesagentur für Arbeit; Kirsten.Singer2@arbeitsagentur.de
Sischka, Benjamin; Ludwig-Maximilians-Universität München; benjamin.sischka@stat.uni-muenchen.de
Söhl, Jakob; Delft University of Technology; j.soehl@tudelft.nl
Söhnlein, Doris; Institut für Arbeitsmarkt- und Berufsforschung; doris.soehnlein@iab.de
Sonnenburg, Anja; Gesellschaft für Wirtschaftliche Strukturforchung; sonnenburg@gws-os.com
Speich, Wolf-Dietmar; Statistisches Landesamt des Freistaates Sachsen; wolf-dietmar.speich@statistik.sachsen.de
Stabenow, Kevin; WWU Münster; kevin.stabenow@wiwi.uni-muenster.de
Steeg, Stefanie; Bundesinstitut für Berufsbildung; steeg@bibb.de
Steger, Almut; Deutsche Bundesbank; almut.steger@t-online.de
Stender, Axel; Stadt Moers; axel.stender@moers.de
Stocker, Toni; Universität Mannheim; stocker@uni-mannheim.de
Stolpe, Susanne; Universitätsklinikum Essen; susanne.stolpe@uk-essen.de
Stöver, Britta; Institut für Statistik an der Leibniz Universität Hannover; stoever@statistik.uni-hannover.de
Strohe, Hans Gerhard; Universitäten Potsdam und Opole; hgstrohe@web.de
Suffert, Martin; Fujitsu TDS GmbH; msuffert@tds.de
Tentscher, Christiane; Statistik der Bundesagentur für Arbeit; christiane.tentscher@arbeitsagentur.de
Teske, Markus; Statistisches Bundesamt; markus.teske@destatis.de
Then, Ralf; Stadt Nürnberg; ralf.then@statistik.nuernberg.de
Thien-Seitz, Uta; Landeshauptstadt München; uta.thien@muenchen.de
Trutzel, Klaus; Amt für Stadtforschung und Statistik; kum.trutzel@t-online.de

van der Leij, Marco; University of Amsterdam; m.j.vanderleij@uva.nl
van der Wurp, Hendrik; TU Dortmund University; vanderwurp@statistik.tu-dortmund.de
Vedder-Stute, Beate; IT.NRW; beate.vedder-stute@it.nrw.de
Vintis, Tim; Universität Bremen; tvintis@uni-bremen.de
Voges, Michelle; Leibniz Universität Hannover; voges@statistik.uni-hannover.de
von der Lippe, Elena; Robert Koch-Institut; vonderlippee@rki.de
von Kalckreuth, Ulf; Deutsche Bundesbank; ulf.von-kalckreuth@bundesbank.de
von Sachs, Rainer; Université catholique de Louvain; rvs@uclouvain.be
Vorgriemler, Daniel; Statistisches Bundesamt; daniel.vorgriemler@destatis.de
Voshage, Ramona; Amt für Statistik Berlin-Brandenburg; ramona.voshage@statistik-bbb.de
Wagner, Ingo; Statistisches Bundesamt; ingo.wagner@destatis.de
Waschipky, Martin; Amt für Statistik und Wahlen Leipzig; martin.waschipky@leipzig.de
Wayand, Jürgen; DStatG; juergen.wayand@arcor.de
Weißbrodt, Thomas; Statistik der Bundesagentur für Arbeit; thomas.weissbrodt@arbeitsagentur.de
Wenger, Kai Rouven; Leibniz University Hannover; wenger@statistik.uni-hannover.de
Wengler, Annelene; Robert Koch-Institut; wenglera@rki.de
Werner, Andrea; Statistik der Bundesagentur für Arbeit; Andrea.Werner7@arbeitsagentur.de
Werth, Stefan; Statistik der Bundesagentur für Arbeit; stefan.werth@arbeitsagentur.de
Willis, Nico; Private; nico.willis@gmx.de
Windmann, Michael; Ludwig-Maximilians-Universität München; michael.windmann@stat.uni-muenchen.de
Winkelmann, Ulrike; Statistisches Landesamt Baden-Württemberg; ulrike.winkelmann@stala.bwl.de
Wittig, Jürgen; Magistrat der Stadt Kassel; juergen.wittig@kassel.de
Wolf, Patrick; Zentrum für Sonnenenergie- und Wasserstoff-Forschung Baden-Württemberg (ZSW);
patrick.wolf@zsw-bw.de
Wolfanger, Tobias; IT.NRW; tobias.wolfanger@it.nrw.de
Wölwer, Anna-Lena; Universität Trier; woelwer@uni-trier.de
Zemann, Christian; Statistik der Bundesagentur für Arbeit; Christian.Zemann@arbeitsagentur.de
Zimmermann, Thomas; Destatis; thomas.zimmermann@destatis.de
Zimmermann, Ralf; Statistik der Bundesagentur für Arbeit; ralf.zimmermann6@arbeitsagentur.de
Zinn, Sabine; SOEP DIW; szinn@diw.de
Zins, Bernhard Stefan; Institut für Arbeitsmarkt- und Berufsforschung; stefan.zins@iab.de
Zwerenz, Karlheinz; Hochschule München; prof.zwerenz@t-online.de
Zwick, Markus; Statistisches Bundesamt; markus.zwick@destatis.de

Building Plans / Gebäudepläne



Gesellschaft braucht Wissenschaft.



Wissenschaft **braucht Daten.**

8| Konferenz für Sozial- und Wirtschaftsdaten

02.–03. März 2020 in Berlin, www.ratswd.de/kswd

Die **8. Konferenz für Sozial- und Wirtschaftsdaten (8|KSWD)** ist das Austauschforum für empirische Wirtschafts- und Sozialforschung. Erhebungs- und Analysemethoden, Datenquellen und Dateninfrastrukturen stehen zur Debatte.

■ **Schwerpunkte der Konferenz**

Evidenzbasierte Politikberatung, FAIRe Registerdaten, Big Data, Datafizierung der Gesellschaft, Kriminalstatistik, Online-Erhebungen, Innovative Zugänge zu sensiblen Daten, Datenerhebung mit Smartphone & Co

■ **Call for Posters**

Prämierte Posterpräsentation und Elevator Pitch für innovative Forschung mit Daten aus den Forschungsdatenzentren



Call for Posters und Anmeldung:
■ www.ratswd.de/kswd



Wahl der Vorschläge

aus der Wissenschaft für die
7. Berufungsperiode (2020–2023) des RatSWD



Jetzt für die Wahl registrieren:
■ www.ratswd.de/wahl

GEFÖRDERT VOM



Bundesministerium
für Bildung
und Forschung



 **Universität Trier**



© Universität Trier

**Alle Informationen zur Statistischen Woche
jetzt auch in der Tagungs-App!**



iPhone



Android



www.statistische-woche.de