Text Analysis & Sentiment Analysis (TASA) with Applications to Finance

Wolfgang Karl Härdle

Cathy Yi-Hsuan Chen
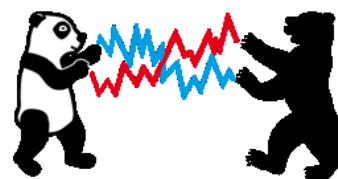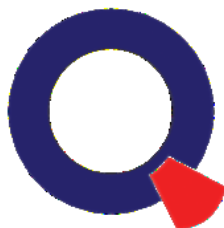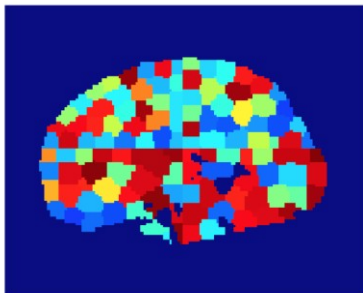
Wolfgang Karl Härdle completed his Dr. rer. nat. in Mathematics at Heidelberg University and received his habilitation in Economics at Friedri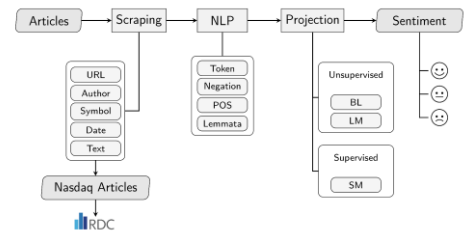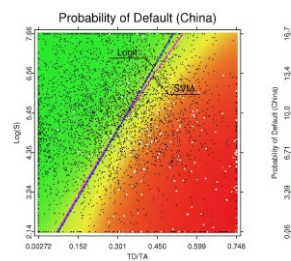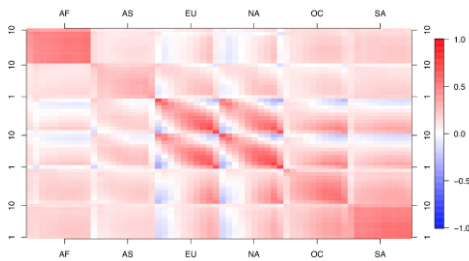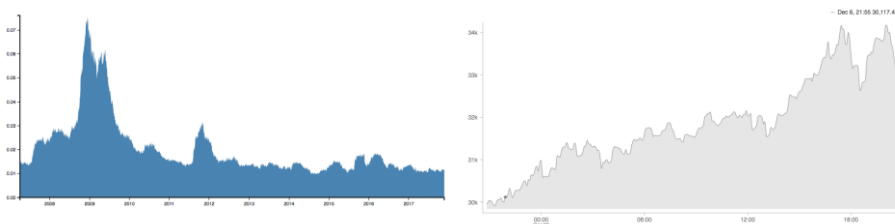ch Wilhelm Universität Bonn. He is now Ladislaus von Bortkiewicz Chair Professor of Statistics at Humboldt-Universität zu Berlin. He was the founder and director of Collaborative Research Center CRC 373 "Quantification and Simulation of Economic Processes" (1994 - 2003) and also the Director of C.A.S.E. (Center for Applied Statistics and Economics) (2001 - 2014). He founded the CRC 649 "Economic Risk" (2005 - 2016) and is now directing the Sino-German International Research Training Group IRTG1792 "High dimensional non stationary time series analysis" (2013-2022). He has proposed the CRIX Crypto Currency Index crix.hu-berlin.de, the Financial Risk Meter FRM frm.hu-berlin.de and QuantNet www.quantlet.de

His research focuses on dimension reduction techniques, computational statistics and quantitative finance. He has published 30+ books and 400 + papers in top statistical, econometrics and finance journals and has a high citation indices across different ranking platforms, see hu.berlin/wkh

He has top notch rankings in REPEC and in other scientometric scales, such as the Handelsblatt ranking, Scopus and Google scholar.

His professional experience includes financial engineering, analysis of unstructured data and dynamic decision analytics. He currently focuses his research on sentiment distillation, crypto currencies and DEDA - Digital Economy & Decision Analytics. He has supervised more than 50 PhD students and is holding up long-term research relations to partners in the USA, Singapore, Prague, Warsaw, Paris, Cambridge, Beijing, Xiamen and Taipei among others.

Cathy Yi-Hsuan Chen

Cathy Yi-Hsuan Chen received her PhD diploma in Finance at National Cheng-Chi University. She is since 2015 Associate Professor at the School of Business & Economics in Humboldt-Universität zu Berlin, and a professor of the International Research Training Group 1792 – High Dimensional Non Stationary Time Series. She is the member of Princeton-Humboldt Cooperation and Collective Cognition Network (CoCCoN), and visiting fellow of Sim Kee Boon Institute for Financial Economics, Singapore Management University.

She has been teaching Master courses at Ladislaus von Bortkiewicz Chair of Statistics at Humboldt-Universität zu Berlin since 2015. The courses she has taught include D*igital Economics & Decision Analytics, Statistics of Financials markets, Quantitative Finance, Economic Risk Seminar*.

She has a five-years of consulting experience in the insurance industry. Through internal rating-based models, developed and customised programming training (Matlab), a company she consulted can benefit from superior risk management, employee training and fulfill all the obligations and regulations required by the authorities. She is currently heading a "transfer project" between Humboldt-Universität and Deutsche Bank, and focussing on credit risk modelling and stress testing.

Her research focuses on quantitative finance, risk modelling and management, behaviour finance and text mining in finance. She has published the papers in reputable journals in the past 10 years. She currently concentrates her research on text extraction, text analytics, textual sentiment distillation, textual sentiment modelling, lexicon construction, investor decision and sentiment.

Some Books
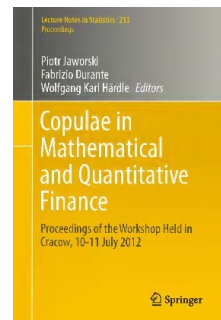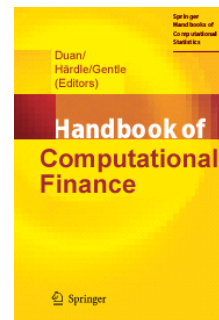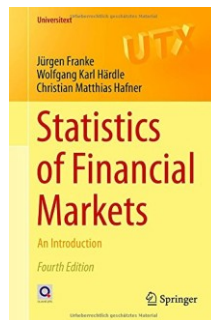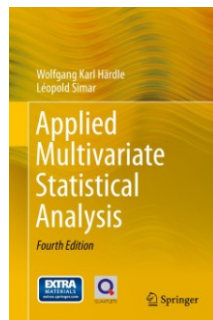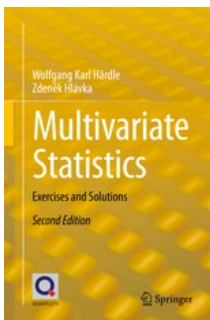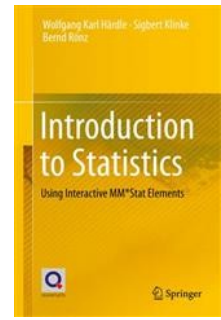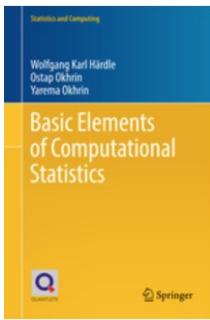
Course Contents

TASA Learning Objectives

Since information mostly exists in language data, the TASA course presents tools and concepts for text data with a strong focus on modeling the econometric effects of language or more specific sentiment. It presents the decision analytics in a way that is understandable for non-mathematicians and practitioners who are confronted with day to day number crunching statistical textual analysis. This course details the development of textual analysis and sentiment projection, and compare the pros and cons of them. The TASA course endows the practitioner with ready to use practical tools for these purposes and applications. All practical examples may be recalculated and modified: software and Quantlets are in www.quantlet.de.

TASA Structure

Data are everywhere and the ubiquitous availability of huge amounts of data makes it necessary to develop smart data analytics. Out of the plethora of tools that are available for many scientific disciplines this course offers for the common data analyst an easy access to all levels of analysis without deep computer programming knowledge. Python is becoming the lingua franca, and can be easily applied for the analysis involved textual data. TASA provides a wide variety of exercises, with Python or R step-by-step demonstrations for web-scraping, Natural Language Processing combined with statistical learning methods.

TASALiterature

Franke J, Härdle WK, Hafner C (2015) Statistics of Financial Markets: an Introduction. 4th ed., Springer Verlag, Heidelberg. ISBN: 978-3-642-54538-2
Chen C YH, Härdle WK, Overbeck L (2017) Applied Quantitative Finance. 3rd extended ed., Springer Verlag, Heidelberg.
Härdle WK, Simar L (2015) Applied Multivariate Statistical Analysis. 4th ed., Springer Verlag, Heidelberg. ISBN 978-3-662-45170-0
Härdle WK, Okhrin O, Okhrin Y (2017) Basics of Computational Statistics, Springer Verlag, Heidelberg.

All examples are presented in R or Python. The Quantlets are available here:

www.quantlet.de

Schedule

Unit 1	
What do we see?	• Basic concepts • Data Management • Structuring Data elements
Unit 2	
Data Analysis	• Natural Language Processing • Stemming, lemmatizing • DTM Dynamic Topic Modeling
Unit 3	
Modern Data Analysis	• Python tools for text mining • Text mining in Quantitative Finance • Applications & Empirics
Unit 4	
Modern Data Analytics	• Cluster Analysis and Classification • Support Vector Machine • CRIX a CRypto currency IndeX
Unit 5	
Sentiment Analysis	• Unsupervised projection: lexicon-based • Supervised projection: sentence-based • News sentiment extraction • Crypocurrency-specific lexicon and sentiment projection
Unit 6	
Smart Data Analytics	• Financial Risk Meter • DDI Networks Topology • Q3 D3 LSA
Unit 7	
Very Smart Data Analytics	• fraud and scam detection • Options on cryptos • Adaptive weight clustering
Unit 8	

We do Smart Data Analytics	• Machine learning in Economics
	• Deep Learning of Forecasts
	• Complexity in Banking, Scores and Networks

Contact

Wolfgang Karl Härdle
Ladislaus von Bortkiewicz Chair of Statistics
School of Business and Economics
Humboldt-Universität zu Berlin
Unter den Linden 6
10099 Berlin, Germany

Telefone   **+49 30 2093 5631**

FAX   **+49 30 2093 5649**

E-Mail   **stat@wiwi.hu-berlin.de**

Links

hu.berlin/93629

hu.berlin/irtg1792

http://crix.hu-berlin.de

hu.berlin/**rdc**

financialriskmeter

hu.berlin/frm

http://quantlet.de