MATT TADDY

Microsoft   taddy@microsoft.com
University of Chicago Booth School of Business   http://taddylab.com

---

PROFESSIONAL POSITIONS

Principal Researcher at Microsoft Research (MSR), 2016-Present
Founder and leader of the MSR ALICE Project on Economic AI

Professor of Econometrics and Statistics, Chicago Booth, 2016-Present (on leave)
Fellow of the Computation Institute, University of Chicago, 2014-Present (on leave)
    Associate Professor, 2012-2016
    Assistant Professor, 2008-2012

Research Fellow at eBay Inc., 2014-2016

EDUCATION

Ph.D. Applied Mathematics and Statistics, June 2008
UNIVERSITY OF CALIFORNIA SANTA CRUZ
Baskin School of Engineering
Advisor: Athanasios Kottas
'Bayesian nonparametric analysis of conditional distributions and inference for Poisson processes'

M.S. Mathematical Statistics, June 2005.
MCGILL UNIVERSITY, MONTRÉAL
Department of Mathematics and Statistics
Advisor: Russell Steele
'Variable selection for neural network treatment classification via parallel sampling'

B.A. Philosophy and Mathematics, June 2003
MCGILL UNIVERSITY, MONTRÉAL

PUBLICATIONS

M. Gentzkow, B. Kelly, and M. Taddy (2017). Text as Data. *NBER working paper 23276.*

J. Hartford, G. Lewis, K. Leyton-Brown, M. Taddy (2017). Counterfactual Prediction with Deep Instrumental Variables Networks. To appear in *ICML 2017.*

S. Upadhyay, K. Chang, M. Taddy, A. Kalai, and J. Zou (2017). Beyond Bilingual: Multi-sense Word Embeddings using Multilingual Context. To appear in *2nd Workshop on representation learning for NLP.*

M. Wan, D. Wang, M. Goldman, M. Taddy, J. Rao, J. Liu, D. Lymberopoulos, and J. McAuley (2017). Modeling consumer preferences and price sensitivities from large-scale grocery shopping transaction logs. *Proceedings of WWW 2017.*

M. Gentzkow, J. Shapiro, and M. Taddy (2016). Measuring polarization in high dimensional data. *NBER working paper 22423.*

M. Taddy (2016). Comment: A regularization scheme on word occurrence rates that improves estimation and interpretation of topical content (Airoldi and Bischof). To appear in the *Journal of the American Statistical Association.*

M. Kolar and M. Taddy (2016). Comment: Coauthorship and citation networks for statisticians (Ji and Jin). To appear in the *Annals of Applied Statistics.*

M. Taddy (2016). One-step estimator paths for concave regularization. To appear in the *Journal of Computational and Graphical Statistics.*

M. Taddy, M. Gardner, L. Chen, and D. Draper (2016). Nonparametric Bayesian analysis of heterogeneous treatment effects in digital experimentation. To appear in the *Journal of Business and Economic Statistics*.

M. Taddy, C. Chen, J. Yun, and M. Wyle (2015). Bayesian and empirical Bayesian forests. Proceedings of the *32nd International Conference on Machine Learning (ICML 2015)*.

M. Taddy (2015). Distributed Multinomial Regression. *Annals of Applied Statistics 9*, 1395–1414.

M. Taddy (2015). Document classification by inversion of distributed language representations. Proceedings of the *53rd Meeting of the Association for Computational Linguistics (ACL 2015)*.

R.G. Gramacy and M. Taddy and S. Tian. Hockey player performance via regularized logistic regression. To appear in the *Handbook of statistical methods for design and analysis in sports*, CRC press.

M. Taddy (2013). Multinomial inverse regression for text analysis. *Journal of the American Statistical Association 108*, 755–770, with discussion.

M. Taddy (2013). Rejoinder: Efficiency and structure in MNIR. *Journal of the American Statistical Association 108*.

R.B. Gramacy and M. Taddy and S.M. Wild (2013). Variable selection and sensitivity analysis via dynamic trees with an application to computer code performance tuning. *Ann. of Applied Stat. 7*, 51–80.

M. Taddy (2013). Measuring political sentiment on Twitter: Factor-optimal design for multinomial inverse regression. *Technometrics 55*, 415–425.

R.B. Gramacy and S.T. Jensen and M. Taddy (2013). Estimating player contribution in hockey with regularized logistic regression. *Journal of Quantitative Analysis of Sports 9*, 97–111.

M. Taddy and A. Kottas (2012). Mixture modeling for marked Poisson processes. *Bayesian Analysis 7*, 335–362.

M. Taddy (2012). On estimation and selection for topic models. Proceedings of the *15th International Conference on Artificial Intelligence and Statistics (AISTATS 2012)*.

M. Taddy, R.B. Gramacy and N.G. Polson (2011). Dynamic trees for learning and design. *Journal of the American Statistical Association 106*, 109–123.

M. Taddy (2010). An auto-regressive mixture model for dynamic spatial Poisson processes: Application to tracking the intensity of violent crime. *Journal of the American Statistical Association 105*, 1403–1417.

M. Taddy and A. Kottas (2010). A Bayesian nonparametric approach to inference for quantile regression. *Journal of Business and Economic Statistics 28*, 357–369.

C.M. Carvalho, H.F. Lopes, N.G. Polson, and M. Taddy (2010). Particle learning for general mixtures. *Bayesian Analysis 5*, 709–740.

H.K.H. Lee, M. Taddy, R.B. Gramacy, and G.A. Gray (2010). Designing and analyzing a circuit device experiment using treed Gaussian processes. *Handbook of Applied Bayesian Analysis, Chap. 28*, Oxford University Press.

R.B. Gramacy and M. Taddy (2010). Categorical inputs, sensitivity analysis, optimization and importance tempering with tgp version 2, an R package for treed Gaussian process models. *Journal of Statistical Software 33*, Issue 6.

H.K.H. Lee, M. Taddy, and G.A. Gray (2010). Selection of a representative sample. *Journal of*

Classification 27, 41–53.

M. Taddy, H.K.H. Lee, G.A. Gray, and J. Griffin (2009). Bayesian guidance for robust pattern search optimization. *Technometrics 51*, 389–401.

M. Taddy and A. Kottas (2009). Markov switching Dirichlet process mixture regression. *Bayesian Analysis 4*, 793–816.

M. Taddy, B. Sansó and H.K.H. Lee (2009). Fast inference for statistical inverse problems. *Inverse Problems 25* 085001.

R.D. Morris, A. Kottas, M. Taddy, R. Furfaro and B.D. Ganapol (2008). A statistical framework for the sensitivity analysis of radiative transfer models. *IEEE Transactions on Geoscience and Remote Sensing 46*, 4062–4074.

G.A. Gray, M. Martinez-Canales, M. Taddy, H.K.H. Lee, and R.B. Gramacy (2006). Enhancing parallel pattern search optimization with a Gaussian process oracle. *Proceedings of the 14th NECDC.*

TEACHING

CHICAGO BOOTH
Big Data 2011+ ('Data Mining' prior to 2014).
Created the course, runs annually for 200 MBAs
See the teaching page on my website for materials
Applied Regression Analysis 2008-2010

2013 NBER ECONOMETRICS LECTURES
Two day course with V. Chernozukov, C. Hansen, M. Gentzkow, and J. Shapiro

TEACHING ASSISTANT at UCSC and McGill in mathematics, applied math, and statistics

SERVICE

Associate Editor at JASA Applications and Case Studies, 2013+

Associate Editor at Statistica Sinica, 2014+

Associate Editor at Bayesian Analysis, 2013+

Co-organizer of the Microsoft Digital Economics Conference, 2016.

Academic Advisory Board member for the National Opinion Research Center (NORC), 2014+

Conference Co-chair, Kilts center conference on Marketing and Big Data at Chicago Booth, 2015+

Academic Board member for Research Computing at the University of Chicago, 2014-2016

Program Chair, ASA section on Statistical Learning and Data Mining, 2014-15

Scientific Comittee member, ISBA 2014

FUNDING AND AWARDS

Emory Williams teaching award (chosen by students), Chicago Booth, 2016.

McKinsey teaching award (chosen by faculty), Chicago Booth, 2016.

Neubauer Family Faculty Fellow, University of Chicago, 2011-2013, 2014-2015

NSF Grant $403,036 over two years, Division of Social and Economic Sciences, September 2016.
Bilateral Bargaining through the Lens of Big Data, with Tadelis, Larsen, and Backus.

Fama-Miller Center Research Grant, 2011-2012
Analysis of Economic and Financial News, with B.T. Kelley and R.S.J. Koijen

Robert L. Graves Scholar, University of Chicago, 2010 - 2011

ISBA Savage Award Honorable Mention, 2010
For outstanding PhD dissertation contribution in applied Bayesian methodology

IBM Corporation Scholar, University of Chicago, 2009 - 2010

Research member of the NMFS Center for Stock Assessment Research, 2007 - 2008

Fonds Québécoise de la recherche sur la nature et les technologies, 2004

Graduate Student Researcher at

    Lawrence Livermore National Laboratory, 2007, advised by Herbie Lee and Bruno Sansó

    NASA Ames, 2007, advised by Athanasios Kottas and Robin Morris

    Sandia National Laboratories, 2006-2007, advised by Genetha Grey and Herbie Lee

    Los Alamos National Laboratory, 2006, advised by Dave Higdon, Herbie Lee, and Bruno Sansó.

    Montréal Jewish General Hospital and URGENCES-SANTÉ, 2004, RA advised by Alain Vandal.

Petro-Canada Olympic Torch scholarship for student athletes, 1999-2000

Canadian 18 & under national sailing champion, men's singlehanded dinghy (laser), 1999

SOFTWARE

R Packages

    distrom: *Distributed multinomial regression*

    gamlr: *Gamma lasso concave penalty regularization paths*

    textir: *Inverse regression for analysis of sentiment in text*

    maptpx: *MAP estimation of latent topic models*

    dynaTree: *Dynamic treed regression and classification* , with R.B. Gramacy

    Bmix: *Sampling algorithms for stick-breaking mixtures*

    tgp: *Nonlinear regression with treed Gaussian processes*, with R.B. Gramacy

Contributor to the gensim python library for natural language processing.

SEMINARS

2016: EC2 Meeting, Toulouse; WU Institute for Statistics and Mathematics, Vienna; CMStatistics, Seville; Harvard CS+Econ seminar; MIT/Harvard Econometrics colloquium; NYU Stern IOMS Department, NYC; NYU Text-as-data seminar, NYC; MSR Digital Economics Conference, NYC; Discussion and invited lecture at the Joint Statistical Meetings (JSM), Chicago IL; ISBA World Meeting, Sardinia; High Dimensional Data Analysis workshop, Fields Institute, Toronto; Bank of England/LBS workshop on fintech, London UK; Becker center, University of Chicago; MIT CSAIL, Cambridge MA; Hockey analytics conference, Simon Fraser University, Vancouver

2015: Columbia University Statistics Department, NYC; University of Illinois at Chicago, Information and Decision Sciences; Big Data workshop, Insper, São Paulo BR; University of Chicago Statistics Department; Duke Statistics, Durham NC; Introductory overview lecture at the Joint Statistical Meetings (JSM), Seattle WA; Applied Topology and Statistics workshop, University of Victoria.; Association for Computational Linguistics (ACL2015), Beijing CN; International Conference on Machine Learning (ICML2015), Lille FR; ISBA Bayesian nonparametrics meeting, Raleigh NC; Conference on economic applications of Big Data, Cambridge-INET, Cambridge UK; UC Santa Cruz, Dept of Applied Math & Stats, Santa Cruz CA; Centre de recherches mathematiques workshop, Montréal QC; UC Berkeley Law and Economics Seminar; Microsoft Research, Cambridge MA; eBay research, San Jose CA; UT Dallas Jindal School of Management

2014: eBay research, San Jose CA; Perspectives on HD data analysis, BIRS, Banff Canada; ISBA Big Data discussion panel, Cancun MX; University of Pennsylvania, Econ Dept, Philadelphia PA; Carnegie Mellon, School of Computer Science, Pittsburgh PA; Stanford, Institute for Research in Social Sciences, Stanford CA; Georgetown, McCourt School of Public Policy, Washington DC; 'Show and Tell' seminar series, UC Research Computing Center, Chicago IL

2013: International Chinese Statistics Association, Hong Kong.; University of Michigan Econ + PoliSci, Ann Arbor MI; Princeton Computer Science, Princeton NJ; eBay Research Labs, San Jose CA; NYU Stern, New York NY; INFORMS Technometrics invited session, Minneapolis MN.; JASA discussion session, JSM, Montréal QC; University of Washington Statistics, Seattle WA; WCBI University of Utah, Alta UT; INFORMS Computing Society Meeting, Santa Fe NM

2012: SAMSI Workshop on Digital Advertising, Durham NC; ISBA, Kyoto, Japan; WU Vienna University of Economics and Business, Austria; Department of Energy CoDA, Santa Fe NM; SBIES, UC Santa Cruz, CA

2011: UC Santa Cruz, Dept of Applied Math & Stats, Santa Cruz CA; Brigham Young University, Statistics Dept, Provo UT; Yeditepe Conference on Bayesian Learning, Istanbul Turkey; AMA Advanced Research Techniques Tutorial, Palm Springs CA; SBIES, Washington University, St Louis MO; University of Toronto Department of Economics; Milton Friedman Institute, University of Chicago

2010: Wharton Statistics, Philadelphia PA; ISBA, Alicante Spain; Cambridge University Statistics Laboratory, Cambridge UK; SBIES, University of Texas, Austin TX; Duke University Department of Statistical Sciences, Durham NC; Johannes Kepler University IFAS, Linz Austria; EBEB X, Rio de Janeiro Brazil

2009: INFORMS, San Diego CA; BISP6, Brixen Italy; Schloss Dagstuhl, Saarland Germany; SAMSI Sequential MC workshop, Durham NC

2006-2008: University of Chicago Graduate School of Business, Chicago IL; London School of Economics, London UK; RAND Corporation, Santa Monica CA; Lawrence Livermore National Laboratory, Livermoretad CA; ICCOPT-II, Hamilton ON; BISP5, Valencia Spain; SIAM-CSE, Costa Mesa CA; SAMSI Computer Experiments Workshop, Vancouver BC