# Daniel J. McDonald

CONTACT
INFORMATION

Department of Statistics
Indiana University
Informatics East 201B
Bloomington, IN 47408

*phone:* (812) 855–7828
*email:* dajmcdon@indiana.edu
*www:* https://dajmcdon.github.io/

RESEARCH
INTERESTS

Machine learning; risk estimation; computational approximations; time series; applications in economics, biology, chemistry, finance and music

EDUCATION

**Carnegie Mellon University**, Pittsburgh, Pennsylvania USA

Doctor of Philosophy in Statistics, May 2012
  **Dissertation:** "Generalization error bounds for time series"
  **Advisors:** Cosma Shalizi and Mark Schervish

Master of Science in Statistics, May 2008

**Indiana University**, Bloomington, Indiana USA

Bachelor of Arts, Economics, summa cum laude, May 2006

Bachelor of Science in Music (Cello) and an Outside Field (Mathematics), magna cum laude, May 2006

ACADEMIC
APPOINTMENTS

**Associate Professor of Statistics** (with Tenure), 2018 –
Indiana University, College of Arts and Sciences, Bloomington
Core faculty, Department of Statistics, Program on Data Science
Adjunct Faculty, Department of Computer Science
Affiliate, Center for Algorithms and Machine Learning

**Visiting Associate Professor of Econometrics and Statistics**, 2018 – 2019
The University of Chicago Booth School of Business

**Assistant Professor of Statistics**, 2012 – 2018
Indiana University, College of Arts and Sciences, Bloomington

**Graduate Research and Teaching Assistant**, 2007 – 2012
Carnegie Mellon University, Department of Statistics, Pittsburgh

AWARDS AND
HONORS

National Science Foundation CAREER Award, 2018

Indiana University Trustees' Teaching Award, 2017

Umesh Gavasakar Memorial Thesis Award, 2012

American Statistical Association, Pittsburgh Chapter Student of the Year, 2012

PEER-REVIEWED
PUBLICATIONS

HOMRIGHAUSEN, D., AND MCDONALD, D.J. (2019+), "Compressed and penalized linear regression," *Journal of Computational and Graphical Statistics*, (in press).

KHODADADI, A., AND MCDONALD, D.J. (2019), "Algorithms for estimating trends in global temperature volatility," in *Proceedings of the 33rd AAAI Conference on Artificial Intelligence (AAAI-19)*, eds. P. V. Hentenryck and Z.-H. Zhou, Association for the Advancement of Artificial Intelligence.

HOMRIGHAUSEN, D., AND MCDONALD, D.J. (2018), "A study on tuning parameter selection for the high-dimensional lasso," *Journal of Statistical Computation and Simulation*, **88**, 2865–2892.

DING, L., AND MCDONALD, D.J. (2017), "Predicting phenotypes from microarrays using amplified, initially marginal, eigenvector regression," *Bioinformatics*, **33**(14), i350–i358.

MCDONALD, D.J. (2017), "Minimax Density Estimation for Growing Dimension," in *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS)*, eds. A. Singh and J. Zhu, vol. 54, pp. 194–203, PMLR.

MCDONALD, D.J., SHALIZI, C.R., AND SCHERVISH, M. (2017), "Nonparametric risk bounds for time-series forecasting," *Journal of Machine Learning Research*, **18**(32), 1–40.

HOMRIGHAUSEN, D., AND MCDONALD, D.J. (2017), "Risk consistency of cross-validation for lasso-type procedures," *Statistica Sinica*, **27**(3), 1017–1036.

HOMRIGHAUSEN, D., AND MCDONALD, D.J. (2016), "On the Nyström and column-sampling methods for the approximate principal components analysis of large data sets," *Journal of Computational and Graphical Statistics*, **25**(2), 344–362.

MCDONALD, D.J., SHALIZI, C.R., AND SCHERVISH, M. (2015), "Estimating beta-mixing coefficients via histograms," *Electronic Journal of Statistics*, **9**, 2855–2883.

LOEWENSTEIN, G., KRISHNAMURTI, T., KOPSIC, J., AND MCDONALD, D.J. (2015), "Does increased sexual frequency enhance happiness?" *Journal of Economic Behavior and Organization*, **116**, 206–218.

HOMRIGHAUSEN, D., AND MCDONALD, D.J. (2014), "Leave-one-out cross-validation is risk consistent for lasso," *Machine Learning*, **97**(1-2), 65–78.

HOMRIGHAUSEN, D., AND MCDONALD, D.J. (2013), "The lasso, persistence, and cross-validation," in *Proceedings of the 30th International Conference on Machine Learning (ICML)*, eds. S. Dasgupta and D. McAllester, vol. 28, pp. 1031–1039, PMLR.

JUE, J.J.S., PRESS, M.J., MCDONALD, D.J., VOLPP, K.G., ASCH, D.A., MITRA, N., STANOWSKI, A.C., AND LOEWENSTEIN, G. (2012), "The impact of price discounts and calorie messaging on beverage consumption: A multi-site field study," *Preventive Medicine*, **55**, 629–633.

MCDONALD, D.J., SHALIZI, C.R., AND SCHERVISH, M. (2011), "Estimating beta-mixing coefficients," in *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics (AISTATS)*, eds. G. Gordon, D. Dunson, and M. Dudík, vol. 15, pp. 516–524, PMLR.

MCDONALD, D.J., AND THORNTON, D.L. (2008), "Primer on the mortgage market and mortgage finance," *The Federal Reserve Bank of St. Louis Review*, **90**(1), 31–46.

WORK IN PROGRESS

MCDONALD, D.J., MCBRIDE, M., GU, Y., AND RAPHAEL, C. (2019), "Markov-switching state space models for uncovering musical interpretation," submitted.

DING, L., AND MCDONALD, D.J. (2019), "High-dimensional prediction with sparse principal components," submitted.

MCDONALD, D.J., AND SHARPNACK, J. (2019), "Exponential family trend filtering on grids," in preparation.

McDonald, D.J., and Shalizi, C.R. (2015), "Empirical macroeconomics and DSGE modeling in statistical perspective," in preparation.

McDonald, D.J., and Loewenstein, G. (2016), "Factor analysis for panel data," in preparation.

McDonald, D.J. (2016), "Sparse additive state-space models," in preparation.

**Technical reports**

McDonald, D.J., and Shalizi, C.R. (2017), "Rademacher complexity of stationary sequences," arXiv:1106.0730.

Homrighausen, D., and McDonald, D.J. (2011), "Spectral approximations in machine learning," arXiv:1107.4340.

McDonald, D.J., Shalizi, C.R., and Schervish, M. (2011), "Estimated VC dimension for risk bounds," arXiv:1111.3404.

McDonald, D.J., Shalizi, C.R., and Schervish, M. (2011), "Generalization error bounds for stationary autoregressive models," arXiv:1103.0942.

McDonald, D.J., Loewenstein, G.F., and Kadane, J. (2009), "The behavior of weight-loss study participants in response to incentives," technical report.

**Grants**

CAREER: Calibrating regularization for enhanced statistical inference. PI. National Science Foundation, DMS–1753171, $400,000, 2018 – 2023.

Collaborative research: Statistical and computational efficiency for massive data sets via approximation-regularization. PI. National Science Foundation, DMS–1407439, $89,911, 2014 – 2018.

High-Dimensional Statistics for Macroeconomic Forecasting. Co-PI. (with C. Shalizi as PI), *Institute for New Economic Thinking*, Grant # INO14–00020, $146,142, 2014 – 2016.

Model Complexity and Prediction Error in Macroeconomic Forecasting. (with C. Shalizi as PI), *Institute for New Economic Thinking*, $170,000, 2011 – 2013.

**Invited talks and short courses**

"Trend filtering in exponential families," University of Chicago and Toyota Technological Institute Machine Learning Seminar 2019

"Regularization, optimization, and approximation: The benefits of a convex combination," Texas A&M University, Department of Statistics 2019

"Regularization, optimization, and approximation: The benefits of a convex combination," University of Toronto, Department of Statistics 2019

"Regularization, optimization, and approximation: The benefits of a convex combination," Colorado State University, Department of Statistics 2019

"Regularization, optimization, and approximation: The benefits of a convex combination," University of Pittsburgh, Department of Statistics 2019

"Matrix sketching for alternating direction method of multipliers optimization," Symposium on Statistics and Data Science, 2018

"Statistical implications of (some) computational approximations," University of Virginia, Department of Statistics, 2018

"Compressed and penalized linear regression," Indiana University, Department of Statistics, 2017

"Predicting phenotypes from microarrays using amplified, initially marginal, eigenvector regression," $25^{th}$ Intelligent Systems for Molecular Biology and the $16^{th}$ European Conference on Computational Biology, 2017

"Predicting phenotypes from microarrays using amplified, initially marginal, eigenvector regression," Joint Statistical Meetings, 2017

"Compressed and penalized linear regression," Toyota Technological Institute at Chicago, Machine Learning Seminar, 2017

"Compressed and penalized linear regression," Pontificia Universidad Católica del Perú, Department of Mathematics and Statistics, 2017

"Minimax density estimation," $20^{th}$ International Conference on Artificial Intelligence and Statistics, 2017

"Estimating $\beta$-mixing coefficients with histograms," American Mathematical Society Spring Central Sectional Meeting, Special Session on Dependence in Probability and Statistics, 2017

"Approximation-regularization for analysis of large data sets," University of Louisville, Department of Bioinformatics and Biostatistics, 2016

"Approximation-regularization for analysis of large data sets," University of California, Davis, Department of Statistics, 2016

"Approximation-regularization for analysis of large data sets," Indiana University-Purdue University Indianapolis, Department of Biostatistics, 2016

"Approximation-regularization for analysis of large data sets," National Center for Atmospheric Research, 2016

"Risk estimation for high-dimensional lasso regression," Joint Statistical Meetings, 2016

"Approximate principal components analysis of large data sets," Yale University, Department of Statistics, 2015

"Approximate principal components analysis of large data sets," Indiana University SOIC, Intelligent & Interactive Systems , 2015

"Short course on the predictive viewpoint," Institute for New Economic Thinking Young Scholar Workshop, New York, 2015

"Approximate principal components analysis of large data sets," Joint Statistical Meetings, 2014

"Statistical machine learning with structured data," Indiana University, Department of Statistics, 2014

"Clustering classical music performance," Université Laval, Department of Mathematics and Statistics, 2014

"Short course on the predictive viewpoint," Institute for New Economic Thinking Young Scholar Workshop, Hong Kong, 2013

"Clustering classical music performance," $15^{th}$ IMS New Researchers Conference, 2013

"The lasso, persistence, and cross-validation," $30^{th}$ International Conference on Machine Learning, 2013

"Nonparametric risk bounds for time-series prediction," Indiana University, Department of Economics, 2012

"Nonparametric risk bounds for time-series prediction," Yahoo! Research, 2012

"Nonparametric risk bounds for time-series prediction," George Mason University, Department of Statistics, 2012

"Nonparametric risk bounds for time-series prediction," Bocconi University, Department of Decision Sciences, 2012

"Nonparametric risk bounds for time-series prediction," Indiana University, Department of Statistics, 2012

"Estimating beta-mixing coefficients," $14^{th}$ International Conference on Artificial Intelligence and Statistics, 2011

"Estimating beta-mixing coefficients," American Statistical Association, Pittsburgh Chapter Annual Meeting, 2011

"Spectral approximation methods: performance evaluations in clustering and classification" The Classification Society Annual Meeting, 2011

"Generalization error bounds for state-space models with an application to economic forecasting," Joint Statistical Meetings, 2010

INDUSTRY
EXPERIENCE

**Alvarez & Marsal**, New York, New York

Statistical Consultant, 2010

Developed methodology to predict membership cessation for a chain of fitness clubs throughout the Northeast.

**Federal Reserve Bank of St. Louis**, St. Louis, Missouri

Research Associate, 2006–2007

Provided research support for 2 economists' research projects intended for publication

PROFESSIONAL
SERVICE

Associate Editor, Reviews
  *Journal of the American Statistical Association*
  *The American Statistician*
Program Committee
  *Conference on Uncertainty in Artificial Intelligence (UAI)*
  *International Conference on Artificial Intelligence and Statistics (AISTATS)*
  *International Conference on Machine Learning and Statistics (ICML)*
Referee
  *Electronic Journal of Statistics, IEEE Transactions on Information Theory, International Conference on Learning Theory (COLT), International Conference on Neural Information Processing Systems (NeurIPS), Journal of the American Statistical Association, Journal of Business and Economic Statistics, Journal of Machine Learning Research, Journal of Optimization Theory and Applications, Journal of Statistical Computation and Simulation, Statistics and Computing, Technometrics*
Grant Reviewer
  *Institute for New Economic Thinking*
  *Division of Mathematics and Statistics, National Science Foundation*
Memberships
  *American Statistical Association, Institute of Mathematical Statistics,*
  *Bernoulli Society*

| | |
|---|---|
| UNIVERSITY SERVICE | Center of Excellence for Women & Technology, REU advisor (2019–present) |
| | Faculty advisor, IU CrossFit Club (2015–present) |
| | Faculty mentor, Faculty Student Mentoring Initiative, Office of the Vice President for Diversity Equity and Multicultural Affairs (2016–present) |
| | Data Science Curriculum Committee, School of Computing, Informatics, and Engineering + College of Arts and Sciences (2014–present) |
| | |
| DEPARTMENT SERVICE | Statistics Department Graduate Studies Committee (2012–2015, 2016–2019) |
| | Statistics Department Colloquium Committee (2012–2018) |
| | Statistics Department Executive Committee (2015–2016) |
| | Statistics Department Undergraduate Studies Committee (2015–2016) |
| | Statistics Department Faculty Hiring Committee (2012, 2013, 2014, 2015, 2018) |
| | Statistics Department Computing Committee (2019–2020) |
| | Statistics Department Committee on Business Statistics (2019–2020) |
| | Chair, Statistics Department Colloquium Committee (2013–2014) |
| | |
| COURSES TAUGHT | Indiana University, 2012–present |
| | STAT–S 301, Introduction to Business Statistics (Fa13, Sp14, Fa14) |
| | STAT–S 432, Applied Linear Models II (Sp16, Sp17, Sp18) |
| | STAT–S 682, Topics in Statistical Machine Learning (Sp14) |
| | STAT–S 721, Advanced Statistical Theory I (Fa12, Fa14) |
| | STAT–S 722, Advanced Statistical Theory II (Sp13) |
| | STAT–S 771, Advanced Data Analysis I (Fa16, Sp17, Fa19) |
| | STAT–S 772, Advanced Data Analysis II (Fa16, Sp17, Fa19) |
| | STAT–S 782, Topics in Statistical Learning Theory (Fa17) |
| | STAT–S 785, Seminar on Statistical Theory (Fa17, Sp17, Fa19) |
| | |
| | University of Chicago Booth School of Business, 2018 |
| | 41911, Advanced Econometrics (Fa18) |
| | |
| | Carnegie Mellon University, 2010–2011 |
| | 36–226, Introduction to Probability and Statistics II (Su10, Su11) |
| | |
| STUDENT ADVISING | Lei Ding (PhD advisor, current) |
| | Yucong Jiang (PhD dissertation committee, IU Computer Science, current) |
| | Chao Tao (PhD dissertation committee, IU Computer Science, current) |
| | Shay Liu (PhD dissertation committee, IU Geology, current) |
| | Zikun Yang (PhD dissertation committee, 2019) |
| | Xuefu Wang (PhD dissertation committee, 2019) |
| | Robert Lunde (PhD dissertation committee, Carnegie Mellon Statistics, 2018) |
| | Arash Khodadadi (MS thesis advisor, 2018) |
| | Rong Jin (PhD dissertation committee, IU Informatics, 2017) |
| | Jia Wang (MS thesis advisor, 2017) |
| | Raksha Kumaraswamy (PhD exam advisor, IU Computer Science, 2017) |
| | Scott Brown (MS thesis committee, 2016) |
| | Sanna Wager (PhD Informatics research rotation, 2015) |
| | Yupeng Gu (PhD dissertation committee, IU Informatics, 2015) |
| | Zikun Yang (PhD exam advisor, 2014) |
| | Lijiang Guo (PhD exam advisor, 2014) |
| | Mikyoung Jun (MS thesis committee, 2013) |