

## Kevin C. Bartz

---

- CONTACT INFORMATION Science Center 906  
Department of Statistics *E-mail:* [bartz@fas.harvard.edu](mailto:bartz@fas.harvard.edu)  
Harvard University *Web:* <http://www.kevinbartz.com>  
1 Oxford St., Cambridge, MA 02138
- RESEARCH INTERESTS Monte Carlo methods, statistical computing, protein structure prediction, random fields, random networks, statistical applications to business, machine learning, data mining
- EDUCATION **Harvard University**, Cambridge, MA  
*Ph.D. candidate, Department of Statistics* **September 2006 to May 2011**  
Dissertation topics:
  - Monte Carlo Maximum Likelihood for Exponential Random Graph Models
  - Estimating Coefficients for Thresholding via the Euler Characteristic Heuristic
  - Fast Loop Sampling Methods for Protein Structure Prediction PredictionAdvisor: [Prof. Samuel C. Kou](#)
- A.M., Statistics* **May 2007**
- California Institute of Technology**, Pasadena, CA  
*B.S., Applied Math; B.S., Economics; with honors* **September 2001 to June 2005**  
Advisor: [Prof. Jeffrey A. Dubin](#)
- PAPERS Kevin Bartz, Joseph Blitzstein and Jun Liu. [Monte Carlo Maximum Likelihood for Exponential Random Graph Models: From Snowballs to Umbrella Densities](#). Submitted to *Social Networks*.
- Kevin Bartz and Dave Kane. [Matching Portfolios](#). Submitted to *Journal of Business and Economic Statistics*.
- Nicola Fuchs-Schündeln and Kevin Bartz. [The Role of Borders, Languages, and Currencies as Obstacles to Labor Market Integration](#). Submitted to *European Economic Review*.
- Charles Nunn, Peter Thrall, Kevin Bartz, Tirthankar Dasgupta and Christopher Boesche. [Do Transmission Mechanisms or Social Systems Drive Cultural Dynamics in Socially Structured Populations?](#) *Animal Behavior*, 2009, volume 77, number 6, pp. 1515-1524.
- Rosie Jones, Kevin Bartz and Benjamin Rey. [Automatically Generating Related Queries in Japanese](#). *Language Resources and Evaluation*, December 2006, volume 40, numbers 3-4, pp. 219-232.
- PAPERS IN PREPARATION Kevin Bartz, Jinfeng Zhang and Samuel Kou. [Fast Loop Sampling Methods for Protein Structure Prediction](#). (estimated 2011)
- Kevin Bartz, Samuel Kou and Robert Adler. [Estimating Coefficients for Thresholding via the Euler Characteristic Heuristic](#). (estimated 2011)
- CONFERENCE PAPERS Kevin Bartz, Jack W. Stokes, John C. Platt, Ryan Kivett, David Grant, Silviu Calinoiu and Gretchen Loihle. [Finding Similar Failures Using Callstack Similarity](#). *Proceedings of the 3rd Workshop on Tackling Computer Systems Problems with Machine Learning Techniques*, 2008.
- Kevin Bartz, Cory Barr and Adil Aijaz. [Natural Language Generation for Sponsored Search Advertisements](#). *Proceedings of the 9th ACM Conference on Electronic Commerce*, 2008, pp. 1-9.

Kevin Bartz, Vijay Murthi and Shaji Sebastian. [Logistic Regression and Collaborative Filtering for Sponsored Search Term Recommendation](#). *Proceedings of the Second Workshop on Sponsored Search Auctions*, 2006.

Kevin Bartz, Samuel Wong, Daniel Fernandez, Jun S. Liu, Samuel C. Kou and Jinfeng Zhang. [FRESS: Fast Fragment Regrowth for Protein Structure Simulation](#). Poster at CASP9, 2010.

Nicola Fuchs-Schündeln, Rima Izem and Kevin Bartz. [Space-time Modeling and Boundary Analysis](#). Poster at the Joint Statistical Meetings, 2009.

RESEARCH  
EXPERIENCE

**Harvard University**, Cambridge, MA

*Research Fellow*

**June 2007 to present**

- Led efforts to build a protein folding system for tertiary structure prediction, [MUSICS](#), for the CASP9 protein folding contest. Collaborated with other graduate students and [Prof. Samuel C. Kou](#), [Prof. Jun S. Liu](#) and [Prof. Jinfeng Zhang](#).
- Implemented and help developed a method to estimate thresholds for hypothesis testing in smooth spatial data. Collaborated with [Prof. Samuel C. Kou](#) and [Prof. Robert J. Adler](#).
- Researched and programmed more efficient estimation methods for exponential random graph models. Collaborated with [Prof. Joseph K. Blitzstein](#) and [Prof. Jun S. Liu](#).
- Developed a spatiotemporal autoregressive model to identify cross-border and within-border correlations in European unemployment rates since 1983. Collaborated with [Prof. Nicola Fuchs-Schündeln](#) and [Prof. Rima Izem](#).

**Microsoft Research**, Redmond, WA

*Research Intern, Knowledge Tools*

**June 2008 to September 2008**

Developed a machine-learned callstack similarity metric to identify similar failures. Deployed a failure search engine based on this metric for Microsoft-wide developer use. Published method and findings at OSDI conference. Worked with [Dr. John C. Platt](#).

**Google**, Mountain View, CA

*Software Engineering Intern, Platforms*

**June 2007 to September 2007**

Assisted in statistical analysis of DRAM errors in the wild, using sampling methods and regularized regression. Acknowledged in 2009 SigMetrics paper.

**Yahoo**, Burbank, CA

*Senior Research Engineer, Search Marketing*

**June 2005 to September 2006**

- Developed machine-learned algorithms to find semantically identical queries, applied to English, Japanese and Korean.
- Researched and deployed product automating generation of grammatical ad copy based on a trained statistical model.
- Applied machine learning algorithms on very large data sets to recommend search terms to advertisers.

**Loyalty Matrix**, San Francisco, CA

*Data Analyst Intern*

**June 2004 to September 2004**

Applied machine learning methods to marketing problems, identifying high-value customer segments using random forest and market basket methods.

TEACHING  
EXPERIENCE

**Harvard University**, Cambridge, MA

*Teaching Fellow*

**September 2007 to present**

- [Statistics 110](#): Introduction to Probability (Fall 2007, 2008, 2009, 2010)

- [Statistics 111](#): Introduction to Theoretical Statistics (Spring 2008, 2009, 2010)  
Performed two weekly one-hour sections, graded problem sets and produced a [collection](#) of almost 100 supplementary course handouts, review problems and solutions. Won teaching distinctions in five of six semesters.

OTHER  
PROFESSIONAL  
EXPERIENCE

**Insightful**, Seattle, WA

*Statistical Programmer Intern*

**June 2003 to September 2003**

Wrote **S+DiscreteChoice**, an S-Plus package implementing several variants of conditional discrete choice models.

**Geode Capital Management**, Boston, MA

*Quantitative Analyst Intern*

**Summers 2001, 2002**

- Wrote real-time performance attribution application.
- Developed backtesting software for long-short investment strategies.

**Consulting and other relevant experience**

**June 2001 to present**

- Consulted for a local hedge fund, developing “matching portfolios” customized benchmarks for equity portfolios using matching techniques from the Rubin Causal Model.
- Participated in 2010 Harvard Business School [Business Plan Contest](#). Developed plan to provide machine-learned box office gross predictions for movie theaters.
- Consulted twice for [AdKnowledge](#), writing algorithms to rank sponsored search ads by projected revenue using generalized linear mixed models.
- Consulted for [World Business Credit](#), designing yield tranches and analyzing risk for A/B commercial mortgage-backed securities.

PATENTS

Kevin Bartz, Vijay Murthi, Benjamin Rey and Shaji Sebastian. [System and method for determining semantically related terms based on sequences of search queries](#). Patent number 7,814,086.

Rosie Jones, Kevin Bartz and Benjamin Rey. [System and method for identifying related queries for languages with multiple writing systems](#). Patent number 7,689,554.

Kevin Bartz, Jay Stokes, John Platt, Ryan Kivett, Silviu Calinoiu, David Grant and Gretchen Loehle. [Similarity detection for error reports](#). Application number 20110066908.

Kevin Bartz, Vijay Murthi and Shaji Sebastian. [System and method for determining semantically related terms](#). Application numbers 20090037399, 20080243826, 20080243480, 20080120072, 20070027865.

Cory Barr and Kevin Bartz. [System and method for predicting a casing variation of a term](#). Application number 20080109274.

Kevin Bartz and Cory Barr. [System and method for predicting a displayable form of a term](#). Application number 20080109273.

HONORS AND  
AWARDS

**Certificate of Distinction for Excellence in Teaching**

*Spring 2008, Fall 2008, Spring 2009, Fall 2009, Spring 2010*

Awarded by the Derek Bok Center, Harvard University. Certificates are given each semester, based on high CUE teaching evaluation scores above 4.5 out of 5 from students.

**Post-Qualifying Talk Award**

*Fall 2008*

Awarded each semester to two post-qualifying talks given in the department. Provides funding to attend a conference of the recipient’s choice.

- PROFESSIONAL SERVICES
- Reviewed three papers for ICML 2009 (International Conference on Machine Learning)
  - Reviewed five papers for NIPS 2009 (Neural Information and Processing Systems)
- LECTURES AND PRESENTATIONS
- “[Sampling Methods for Fragment Regrowth in Protein Structure Prediction.](#)” Harvard University Post-qualifying Talk, December 2009. Florida State University Statistics Seminar, May 2010, Tallahassee, FL. Harvard Statistics 321 course, March 2011.
  - “[Statistical Energy for Protein Structure Prediction.](#)” Harvard University Post-qualifying Talk, May 2010. Harvard Statistics 115 course, March 2010.
  - “[Graphs, Bridges and Snowballs: Monte Carlo Maximum Likelihood For Exponential Random Graph Models.](#)” Harvard University Post-qualifying Talk, January 2009. Harvard Statistics 385 course, April 2009. Harvard Statistics 221 course, May 2007.
  - “[A Fast Algorithm to Detect Signals in Random Fields.](#)” Workshop on Random Fields and Stochastic Geometry, February 2009, Banff, AB, Canada. Harvard University Post-qualifying Talk, May 2009. Harvard University Statistics 218 course, May 2008.
  - “[Matching Portfolios.](#)” Harvard University Statistics 370 course, March 2009. Harvard University Post-qualifying Talk, December 2010 (expected).
  - “[Space-time Modeling and Boundary Analysis.](#)” Joint Statistical Meetings, August 2009, Washington, DC. Harvard University Workshop on Spatial Modeling, May 2008. Harvard University Statistics 225 course, January 2008.
  - “[Pairwise Copula Models.](#)” Harvard University Statistics 370 course, April 2009.
  - “[Finding Similar Failures With Callstack Similarity.](#)” Workshop on Tackling Computer Systems Problems with Machine Learning Techniques, December 2008, San Diego, CA.
  - “[Kernel Estimators for Univariate Binary Regression.](#)” Harvard University Statistics 325 course, December 2007.
  - “[Continuous Treatments in Causal Inference.](#)” Harvard University Statistics 233 course, December 2007.
  - “[Logistic Regression and Collaborative Filtering for Sponsored Search Term Recommendation.](#)” Workshop on Sponsored Search Auctions, March 2006, Ann Arbor, MI.
- COMPUTING SKILLS
- Statistical packages: R, S-Plus; C++ mathematical libraries (e.g., LAPACK and BLAS).
  - Languages: C++, C#, Perl.
  - Databases: MySQL, SQLite, PostgreSQL, BerkeleyDB.
  - Parallel computing: Harvard HPC cluster, Amazon EC2.
- REFERENCES
- Prof. Samuel C. Kou, Advisor, Harvard Statistics, [kou@stat.harvard.edu](mailto:kou@stat.harvard.edu)
  - Prof. Robert J. Adler, Technion-Israel Electrical Engineering, [robert@ieadler.technion.ac.il](mailto:robert@ieadler.technion.ac.il)
  - Prof. Jun S. Liu, Harvard Statistics, [jliu@stat.harvard.edu](mailto:jliu@stat.harvard.edu)
  - Prof. Joseph K. Blitzstein, Harvard Statistics, [blitz@fas.harvard.edu](mailto:blitz@fas.harvard.edu)
  - Dr. John C. Platt, Microsoft Research, [John.Platt@microsoft.com](mailto:John.Platt@microsoft.com)
  - Prof. Nicola Fuchs-Schündeln, University of Frankfurt Finance, [fuchs@wiwi.uni-frankfurt.de](mailto:fuchs@wiwi.uni-frankfurt.de)