

**Blakeley B. McShane**  
[mcsbane@wharton.upenn.edu](mailto:mcsbane@wharton.upenn.edu)  
215-327-5512

**Work Address:**

Department of Statistics  
The Wharton School  
University of Pennsylvania  
400 Jon M. Huntsman Hall  
3730 Walnut Street  
Philadelphia, PA 19104-6340

**Home Address:**

1829 Pine Street  
Apartment 305  
Philadelphia, PA 19103

**Education:**

University of Pennsylvania  
The Wharton School

Philadelphia, PA

Ph.D. in Statistics, May 2010 (Anticipated)

Thesis: Integrating Machine Learning Methods with Hidden Markov Models: A New Approach to Categorical Time Series Analysis with Application to Sleep Data

Thesis Advisor: Abraham Wyner, Department of Statistics

Marketing Advisor: Eric Bradlow, Department of Marketing

University of Oxford  
Studies in Philosophy, October 2004 – June 2005

Oxford, UK

University of Pennsylvania  
The College of Arts and Sciences

Philadelphia, PA

M.A. in Mathematics, May 2003

Thesis: Mathematical Methods in Pricing Rainbow Options

Thesis Advisor: Jonathan Block, Department of Mathematics

B.A. in Mathematics, May 2003, *Summa Cum Laude, Distinction in Mathematics*

The Wharton School

B.S. in Economics, May 2003, *Summa Cum Laude*

**Refereed Publications:**

Kiser, R., Asher, M., and McShane, B. B. (2008), "Let's Not Make a Deal: An Empirical Study of Decision Making in Unsuccessful Settlement Negotiations." *Journal of Empirical Legal Studies*, Vol. 5, No. 3, 551-591.

McShane, B., Adrian, M., Bradlow, E.T., and Fader, P.S. (2008), "Count Models Based on Weibull Interarrival Times." *Journal of Business and Economic Statistics*, Vol. 26, No. 3, 369-378.

Jensen, S.T., McShane, B., and Wyner, A.J. (forthcoming as a discussion paper), "Hierarchical Bayesian Modeling of Hitting Performance in Baseball." *Bayesian Analysis*.

**Invited Publications:**

McShane, B., Jensen, S.T., and Wyner, A.J. (forthcoming), “Statistical Learning Methods for Modeling Sleep in Mice.” *Transactions of the Deming Conference on Applied Statistics*.  
**(Thesis-related Paper)**

**Submitted Papers:**

McShane, B., Wyner, A.J., Jensen, S.T., Pack, A., and Galante, R. “Characterization of the Bout Durations of Sleep and Wakefulness: New Metrics for Summarizing Sleep.” Submitted to *Sleep*.  
**(Thesis-related Paper)**

McShane, B., Braunstein A., Piette, J. and Jensen, S. “A Bayesian Variable Selection Approach to Major League Baseball Hitting Metrics.” Submitted to the *Journal of the Royal Statistical Society, Series C*.

**Papers in Progress:**

McShane, B., Bradlow, E., and Berger, J. “Multivariate Visual Diffusion for Social Groups: How Social Identity Influences When and What People Buy.” **(Job Market Paper)**

McShane, B., Jensen, S.T., and Wyner, A.J. “Integrating Machine Learning Methods with Hidden Markov Models: A New Approach to Categorical Time Series Analysis.” **(Thesis-related Paper)**

McShane, B. and Wyner, A.J. “A Statistical Analysis of Multiple Temperature Proxies: Are Reconstructions of Surface Temperatures Over the Last 1000 Years Reliable?”

McShane, B. “More Evidence Contrary to the Statistical View of Boosting.”

Wyner, A. and McShane, B. “Propensity Score Estimation with Machine Learning Methods: What are the Risks of Overfitting?”

Asher, M., Baker T., and McShane, B. “A Quantitative Review of Securities Fraud Class Action Lawsuits: Do the Merits Matter?”

**Conference Presentations:**

“Statistical Learning Methods for Modeling Sleep in Mice.” To be presented December 2009 at the Deming Conference on Applied Statistics, Atlantic City, NJ.

“Evaluating Baseball Metrics Using a Point-Mass Mixture Random Effects Model.” To be presented September 2009 at the New England Symposium on Statistics in Sports, Boston, MA.

“Social Identity and Multivariate Visual Diffusion.” Presented June 2009 at INFORMS Marketing Science Conference 2009, Ann Arbor, MI.

“New Metrics for Summarizing Sleep Bout Durations in Mice.” Presented May 2009 at Center for Sleep and Respiratory Neurobiology Research Retreat, Philadelphia, PA.

“Are Reconstructions of Surface Temperatures Over the Last 1000 Years Reliable?” Presented February 2009 at Information Theory and Applications Workshop, San Diego, CA.

“A Multivariate Diffusion Model for Social Groups: How Social Identity Influences When and What People Buy.” Presented October 2008 at INFORMS 2008 Annual Meeting, Washington, D.C.

“Statistical Prediction of Sleep in Mice Based on Video Data.” Presented May 2008 at Center for Sleep and Respiratory Neurobiology Research Retreat, Philadelphia, PA.

### **Teaching:**

Statistics 541: Advanced Statistical Methods Spring 2007  
*Teaching Assistant* (to Professor Andreas Buja)

First-year doctoral course designed to teach computational methods including the bootstrap, bagging, cross-validation, CART, non-parametric function estimation, and other techniques.

Statistics 111: Introductory Statistics Fall 2006, Spring 2006

*Teaching Assistant* (to Professors Shane Jensen and Abraham Wyner)

Undergraduate course covering probability, confidence intervals, hypothesis tests, and linear regression.

Statistics 611/612: MBA Mathematics Review Summer 2006

*Teaching Assistant* (to Professor Abraham Wyner)

Review course for incoming MBA students covering algebra, differential calculus, and integral calculus.

### **Employment:**

Invite Media, Philadelphia, PA Sept. 2007 – Oct. 2008

*Vice President of Statistics*

- Built an ad-serving engine to perform targeted banner advertising using click-through and social network data.
- Created a yield manager which prices CPC and CPA campaigns in CPM terms thereby enabling the managers of these campaigns to bid on ad exchanges.
- Advised the executive team on issues at the intersection of strategy and technology.

D. E. Shaw & Company, New York, NY Summer 2007

*Quantitative Analyst*

- Developed a suite of software tools to perform Markov Chain Monte Carlo methods for posterior sampling and integration, Bayesian regression, principal components analysis, and various re-sampling techniques.
- Began developing a risk model for the firm’s convertible bond strategy to assess the risk approximate price of bonds in the portfolio which traded infrequently using data from frequently-traded bonds and corresponding equity prices.

George Weiss Associates, New York, NY Jul. 2003 – Sept. 2004

*Quantitative Strategies Analyst*

Summer 2002

- Designed and implemented a quantitative model which traded equities in the financial sector profitably (Sharpe Ratio of 1.97) and held responsibility for managing the daily trade list.
- Invented a methodology to account for geopolitical risk in the fund's portfolios, both as a tool to assess when such factors affected our portfolios and to hedge out such risk.
- Designed historical backtests to assess the profitability of various rule-based trading criteria.
- Created a program to be run daily to alert traders of aberrant stock movements over the preceding weeks.
- Conducted a review of various models already in use by the firm.

McKinsey & Company, New York, NY

Summer 2001

*Business Analyst*

- Staffed in the pharmaceutical industry on a project team responsible for the strategic positioning of a new drug.
- Assisted in creation of a metrics dashboard, senior management's primary tool for measuring the drug's success.
- Conducted a comprehensive review of the client's top competition and briefed senior McKinsey personnel in preparation for meetings with the client's management team.

#### **Grants:**

NIH/SBIR Training Grant for Sleep Research, July 2006 – Present.

Baker Retailing Initiative Research Grant, 2009.

#### **Honors:**

W. Edwards Deming Graduate Student Scholar, 2009

Thouron Award (full academic scholarship to study in the United Kingdom)

Phi Beta Kappa

Benjamin Franklin Scholar

Joseph Wharton Scholar

Dean's List 1998-1999, 1999-2000, 2000-2001, 2001-2002, 2002-2003

William D. Gordon Award (given to graduating Wharton undergraduate with highest GPA)

Beta Gamma Sigma

Golden Key Society

#### **Affiliations:**

American Marketing Association

Institute for Mathematical Statistics

#### **Graduate-level Courses:**

Analysis (1 year)

Abstract Algebra (1 year)

Geometry-Topology, Differential Geometry (1 year)

Applied Probability Models for Marketing

Applied Stochastic Processes

Probability & Stochastic Processes (1 year)

Mathematical Statistics

Bayesian Methods & Computation

Steve Shatz

Antonella Grassi

Ulrich Christ

Peter Fader

Maria Rieders

Robin Pemantle

Dylan Small

Shane Jensen

Linear Statistical Models  
Applied Statistical Methods  
Data Analysis in Marketing  
Advanced Mathematical Statistics: Asymptotics  
Stochastic Calculus & Financial Applications  
Shrinkage Methods in Statistics  
Inductive Statistical Methods  
Financial and Economic Time Series  
Bayesian Statistical Theory & Methods  
Advanced Probability: Probability Inequalities &  
Machine Learned  
Experimental Design & Observational Studies  
Mathematical Models in Marketing  
Infinite Gaussian Sequence Models  
Machine Learning  
Sample Survey Methods  
Forecasting & Time Series  
Independent Studies:  
Markov & Renewal Processes with Applications  
Machine Learning with Application to Neuroscience  
Diffusion Models in Marketing  
Information Theory

**Seminars Attended:**

Statistics Department Colloquium  
Statistics Department Graduate Student Colloquium  
Marketing Department Colloquium  
Marketing Department Graduate Lunch Seminar

**References:**

Eric T. Bradlow\*  
The K. P. Chao Professor  
Professor of Marketing, Statistics, and Education  
The Wharton School  
University of Pennsylvania  
700 Jon M. Huntsman Hall  
3730 Walnut Street  
Philadelphia, PA 19104  
(215) 898-8255

Edward I. George\*  
Universal Furniture Professor  
Professor of Statistics  
The Wharton School  
University of Pennsylvania  
400 Jon M. Huntsman Hall

Don Morrison  
Andreas Buja  
Eric Bradlow  
Mark Low  
J. Michael Steele  
Lawrence Brown  
Richard Berk  
J. Michael Steele  
Linda Zhao  
J. Michael Steele  
  
Dylan Small  
Christophe Van den Bulte  
Mark Low  
Mikhail Traskin  
Dylan Small  
Robert Stine  
  
Shane Jensen  
Abraham Wyner  
Eric Bradlow  
Abraham Wyner

September 2005 – Present  
September 2005 – Present  
January 2008 – Present  
January 2008 – Present

Peter S. Fader\*  
Francis & Pei-Yuan Chia Professor  
Professor of Marketing  
The Wharton School  
University of Pennsylvania  
700 Jon M. Huntsman Hall  
3730 Walnut Street  
Philadelphia, PA 19104  
(215) 898-1132

Shane T. Jensen  
Assistant Professor of Statistics  
The Wharton School  
University of Pennsylvania  
400 Jon M. Huntsman Hall  
3730 Walnut Street

3730 Walnut Street  
Philadelphia, PA 19104  
(215) 898-8229

Philadelphia, PA 19104  
(215) 573-2211

Abba M. Krieger\*  
Robert Steinberg Professor  
Professor of Statistics, Operations & Information  
Management, and Marketing  
The Wharton School  
University of Pennsylvania  
400 Jon M. Huntsman Hall  
3730 Walnut Street  
Philadelphia, PA 19104  
(215) 898-6805

Abraham J. Wyner  
Associate Professor of Statistics  
The Wharton School  
University of Pennsylvania  
400 Jon M. Huntsman Hall  
3730 Walnut Street  
Philadelphia, PA 19104  
(215) 898-2439

\* letter-writers

---

### **Paper Abstracts:**

#### **Refereed Publications:**

Kiser, R., Asher, M., and McShane, B. B. (2008), "Let's Not Make a Deal: An Empirical Study of Decision Making in Unsuccessful Settlement Negotiations." *Journal of Empirical Legal Studies*, Vol. 5, No. 3, 551-591.

We evaluate a large quantitative dataset of civil litigation cases, compare the ultimate award or verdict to the parties' pre-trial settlement positions, and apply behavioral economics theories to understand the results. This paper grew from work I did as an undergraduate.

McShane, B., Adrian, M., Bradlow, E.T., and Fader, P.S. (2008), "Count Models Based on Weibull Interarrival Times." *Journal of Business and Economic Statistics*, Vol. 26, No. 3, 369-378.

We derive a generalized model for count data assuming Weibull interarrival times, overcoming an analytically intractable problem by using a polynomial expansion. We then introduce gamma heterogeneity and show how this model can model both over- and underdispersed count data, allow covariates to be introduced in a straightforward manner through the hazard function, and be computed in standard software.

Jensen, S.T., McShane, B., and Wyner, A.J. (forthcoming as a discussion paper), "Hierarchical Bayesian Modeling of Hitting Performance in Baseball." *Bayesian Analysis*.

We build a sophisticated Bayesian hierarchical model to predict the hitting performance of Major League Baseball players based on past performance and covariates such as age and position. We compare the forecasts of our model to current sabermetric methods on the 2006 season and discuss successes and limitations.

### **Invited Papers:**

McShane, B., Jensen, S.T., and Wyner, A.J. “Statistical Learning Methods for Modeling Sleep in Mice.” *Transactions of the Deming Conference on Applied Statistics* (forthcoming). **(Thesis-related Paper)**

We develop a new statistical methodology to determine the sleep behavior of mice based on video data, allowing researchers access to high throughput, instantaneous classification of sleep states. The methodology builds upon recent advances in machine learning, specifically Random Forests, combined with Hidden Markov Models to account for local time-dependence.

### **Submitted Papers:**

McShane, B., Wyner, A.J., Jensen, S.T., Pack, A., and Galante, R. “Characterization of the Bout Durations of Sleep and Wakefulness: New Metrics for Summarizing Sleep.” Submitted to *Sleep*. **(Thesis-related Paper)**

The standard summary statistics used by sleep researchers to assess sleep have no power to differentiate between different strains of mice despite scientific researchers knowledge that the strains vary quite considerably in underlying sleep behavior. We propose a mathematical model that includes knowledge of the previous sleep state and discriminates between “short” and “long” sleep bouts. From this model, we derive a number of sufficient statistics which show large differences between various strains of mice.

McShane, B., Braunstein A., Piette, J. and Jensen, S. “A Bayesian Variable Selection Approach to Major League Baseball Hitting Metrics.” Submitted to the *Journal of the Royal Statistical Society, Series C*.

The vast proliferation of data and analyses over the past decade has raised a concomitant problem: with so many variables being tracked on individuals throughout time, which ones demonstrate signal and which are noise? We adapt existing variable selection approaches for this particular problem and use Major League Baseball hitting data as a case study. With the rise of *Moneyball* and sabermetrics, traditional measures of hitting performance (*e.g.*, AVG, RBIs) have been supplanted by new measures (*e.g.*, OBA, ISO). We find several of these measures contain no signal and that those that do contain signal do not add substantive information because they are highly correlated with traditional measures of power, speed, ability to make contact, and plate discipline.

### **Papers in Progress:**

McShane, B., Bradlow, E., and Berger, J. “Multivariate Visual Diffusion for Social Groups: How Social Identity Influences When and What People Buy.” **(Job Market Paper)**

How does the type of people who own a particular product influence others’ likelihood of purchase? How do these effects vary across products? We are interested in the role social identity plays in product adoption with an application to US automobile sales data. We use probability models to identify the relevant social groups as well as whether the adoption of a given make and model by one social group can be explained by previous adoption by another social group. In particular, we build multivariate diffusion models where, in addition to using overall adoption as a covariate (as in the Bass model), we also use summary statistics from each of the group’s diffusions. We examine how these

social effects vary across geographic areas as well as makes, models, and categories of cars.

McShane, B., Jensen, S.T., and Wyner, A.J. “Integrating Machine Learning Methods with Hidden Markov Models: A New Approach to Categorical Time Series Analysis.” **(Thesis-related Paper)**

Machine learning methods have proven extremely successful at the task of classification on a wide variety of problems, in large part because these methods focus on minimizing classification error. However, in more difficult classification situations, the performance of these procedures can be significantly reduced. For example, these methods generally assume that there is no time series structure in the data, that is they assume conditional probability of the class label  $Y_t$  for observation  $t$  given its covariates  $X_t$  is conditionally independent of the rest of the data. In such a setting, one may be able to gain additional discriminatory power by modeling dependencies within the observed data beyond the usual i.i.d. assumption of most machine learning procedures. Our method combines conventional classification methods with a hidden Markov model. We extend this result to cover higher-order Markov chains, variable length Markov chains, and generalized Markov chains and we introduce the concept of a transition-dependent generalized Markov chain

McShane, B. and Wyner, A.J. “A Statistical Analysis of Multiple Temperature Proxies: Are Reconstructions of Surface Temperatures Over the Last 1000 Years Reliable?”

We conduct a thorough review and validation of the statistical methodologies employed by climate scientists in order to create paleoclimatic reconstructions (i.e., local or global temperature reconstructions based on proxy data such as tree rings, ice cores, bristlecones, coral reefs, etc.). We also provide our own reconstruction with appropriate standard errors; though our mean forecast does not differ that greatly from those of climate scientists, the standard errors are sufficiently wide as to make drawing conclusions difficult.

McShane, B. “More Evidence Contrary to the Statistical View of Boosting.”

The statistical view of boosting algorithms focuses on optimization and provides a theoretical framework within which it is possible to derive theorems and create new algorithms. However, many unanswered questions still remain. This paper provides examples via simulation which show flaws in suggestions and methods derived from the statistical view.

Wyner, A. and McShane, B. “Propensity Score Estimation with Machine Learning Methods: What are the Risks of Overfitting?”

The propensity score literature contains many examples where using an estimate of the propensity score produces better estimates of treatment effects than using the propensity score itself (when known). Applied researchers have concluded from this that an overfit propensity score model will not harm estimation of treatment effects and is generally preferable to an underfit one. This claim is investigated.

Asher, M., Baker T., and McShane, B. “A Quantitative Review of Securities Fraud Class Action Lawsuits: Do the Merits Matter?”

This paper investigates a quantitative dataset of over 2,500 securities fraud class action lawsuits. Various financial theories are tested and the role of the “merits” of the cases are examined statistically. We find statistically significant predictors of the motion to dismiss and that these predictors do not help much in classification. We also detect subtle signal in settlement amounts leading to a non-trivial degree of predictability.