Guillaume Allaire Pouliot
scholar.harvard.edu/pouliot
pouliot@fas.harvard.edu

HARVARD UNIVERSITY

Placement Director: David Cutler DCUTLER@HARVARD.EDU 617-496-5216
Placement Director: Oliver Hart OHART@HARVARD.EDU 617-496-3461
Graduate Administrator: Brenda Piquet BPIQUET@FAS.HARVARD.EDU 617-495-8927

Office Contact Information:

Littauer Center
1805 Cambridge Street
Cambridge, MA 02138

Undergraduate Studies:

Honors B.A., University of Chicago, June 2010
Thesis Title: "*Compromised Randomization in Abecedarian Preschool Program Experiment*" (Lawrence G. Goldberg Honors Thesis Prize)
Advisor: James J. Heckman

M.S. in Statistics, June 2010 (Concurrent)
Thesis Title: "*Statistical Issues in Visual Field Analysis*"
Advisor: Debashis Mondal

Graduate Studies:

Harvard University, 2010 to present

Ph.D. Candidate in Economics
Thesis Title: "*Applications of Statistical Machine Learning and Spatial Statistics in Economics*"

Expected Completion Date: May 2016

References:
Professor Gary Chamberlain (committee chair)
Harvard University
gary_chamberlain@harvard.edu, (617) 495-1869

Professor Edward Glaeser
Harvard University
eglaeser@harvard.edu, (617) 495-0575

Professor Elie Tamer
Harvard University
elietamer@fas.harvard.edu, (617) 496-1526

Professor Neil Shephard
Harvard University
shephard@fas.harvard.edu, (617) 495-5496

Teaching and Research Fields:

Primary fields: Econometrics and Statistics, Applied Econometrics
Secondary fields: Applied Microeconomics, Statistical Machine Learning, Optimization

Teaching Experience:

Fall 2015	Teaching fellow, Econometric Methods (advanced undergrad), Harvard, for Elie Tamer
Spring 2014	Teaching fellow, Econometric Methods (advanced undergrad), Harvard, for Gary Chamberlain
Spring 2014	Teaching fellow, Econometrics II (grad core sequence), Harvard, for Gary Chamberlain
Fall 2013	Teaching fellow, Introduction to Econometrics, Harvard, for Jeffrey Miron
Spring 2013	Teaching fellow, Econometric Methods (grad), Harvard, for Guido Imbens
Fall 2012	Teaching fellow, Introduction to Econometrics, Harvard, for Eric Chaney
Spring 2012	Teaching fellow, Econometric Methods (grad), Harvard, for Guido Imbens
Spring 2010	Teaching fellow, Intermediate Statistics, University of Chicago, for David Degras
Fall 2009	Teaching fellow, Intermediate Statistics, University of Chicago, for Linda Collins

Research Experience and Other Employment:

Summer 2011	University of Chicago, research assistant for Mathias Drton (University of Chicago, Statistics)
Summer 2009	University of Chicago, research assistant for Azeem Shaikh
2007—2008	University of Chicago, research assistant for James Heckman

Honors, Scholarships, and Fellowships:

2010 to present	Ashford Fellowship
2006—2010	University Scholarship, Phi Beta Kappa

Research:

"*Spatial Econometrics for Misaligned Data*" (Job Market Paper)

Abstract: In economics, many data sets providing the specific location of households, firms, villages, or other economic units are matched by location to data with geographic features such as rainfall in order to study their economic impact. These data sets are often misaligned: the locations of observations in both datasets do not generally coincide. In this article, I study a class of regression problems with spatially correlated variables. This includes regression analysis with misaligned data. I introduce a quasi-maximum likelihood estimator as well as a more robust companion method which does not require specification of the regression error covariance. For both methods, I obtain new central limit theorems for spatial statistics, which are of independent interest. I propose computational strategies and investigate their performance. In simulation, I find that the methods I recommend, along with the asymptotic distribution theory I derive, yield more reliable estimates and confidence intervals than previously recommended methods. In the reanalysis of two datasets, I find that the suggested methods yield quantitatively and qualitatively different conclusions.

"*Equivalence of Multicategory SVM and Simplex Cone SVM*"

Abstract: I show that the multicategory SVM (MSVM) of Lee, Yin, and Wahba (2004) is equivalent to the Simplex Cone SVM of Mroueh, Poggio, Rosasco, and Slotine (2012). Leveraging that result, I obtain finite-dimensional kernel asymptotic distribution theory for multicategory SVM and compare asymptotic covariance matrices to explain analytically the impressive comparative performance of the One-vs-Rest SVM procedure for multicategory data. I display a case in which it is strictly more efficient than MSVM.

"*Simplex Monte Carlo and Fast MCMC Sampling of Bootstrap Distribution*"

Abstract: I develop a Monte Carlo method for sampling the empirical bootstrap distribution of certain robust estimators. Specifically, I suggest using a Metropolis-Hastings algorithm to sample the bootstrap distribution of estimators whose loss function, such as the check or hinge loss, implies a discrete support for the estimated coefficient. This article focuses on the case of quantile regression. In order to construct a suitable proposal distribution, I generalize the dynamics of Hamiltonian Monte Carlo and apply them to sampling the bootstrap distribution of estimators which are the solution of a linear program.

"*Data Mining the Hair in your Soup*" (joint with Mike Luca, Presented at the Joint Statistical Meeting)

Abstract: The city of San Francisco has a limited number of sanitation inspectors and many restaurants to inspect. It is more efficient to send inspectors to restaurants which are more likely to get caught for sanitation norm violations. We take up the task of predicting which restaurants these are. Our dataset consists of a comprehensive list of San Francisco restaurants, their general characteristics, their Yelp reviews (provided by Yelp), as well as past inspection records. Using a supervised latent Dirichlet allocation (LDA) topic model, we improve prediction accuracy using the information from Yelp reviews, and produce interpretable natural language processing (NLP) output. We study the comparative performance of inference output using Gibbs sampling and variational inference.

"*Existence of Maximum Likelihood for Directed Acyclic Graphs*" (with Mathias Drton and Christopher Fox)

Abstract: In this paper, we give the necessary and sufficient number of observations required for the Gaussian maximum likelihood estimator of any directed (acyclic or cyclic) graphical model to be bounded.

"*Latent subcategories for support vector machines*" (with Jules Marchand-Gagnon)

Abstract: One way to accommodate very different feature descriptions for the same category is to use a flexible kernel. However, if the differences in features within a category are well divided into few clusters of the features, allowing for latent subcategories can yield a good fit while avoiding issues of overfitting. We suggest a semi-supervised approach, latent subcategory SVM (LS^2VM), which learns unlabeled subcategories within labeled categories, allowing for accurate classification using more categories but less flexible (even linear) kernels. The resulting optimization problem is a mixed-integer problem, but is exactly equivalent to a difference-convex program. We give a fast algorithm for implementation, and compare performance with Bayesian approaches.

"Probabilistic SVM and Fast MCMC Sampling of Bootstrap Distributions"

Abstract: I show that, when using multicategory support vector machines, more accurate out-of-sample predictions may be obtained by using a decision rule which accounts for the variability of the separating hyperplane. In order to evaluate such a decision rule, I suggest a fast procedure which relies on a variant of the Hamiltonian Monte Carlo algorithm to sample the bootstrap distribution of the multicategory support vector machine decision function. In simulations and applications on benchmark datasets, I find that the precision of predictions can improve substantially, especially for points distant from the bulk of the training set.