Optimization-Conscious Econometrics

ECON 2148

Harvard University, Autumn 2020

Instructor: Guillaume A. Pouliot
Time: Tuesday and Thursday, 3:00-4:15
Teaching Assistant: TBD
Section Time: TBD

Course Description

Modern research in econometrics often intersects with machine learning and big data. Likewise, while the overlap of econometrics with optimization and operations research has traditionally been limited, previously intractable large scale or combinatorially difficult econometrics problems are now being solved using modern optimization software and heuristics.

This lays out a rich research agenda and opens up consequential new questions for econometricians. How can machine learning methods be used for econometric regression analysis and causal inference? How can modern optimization methods be applied to solve previously intractable econometric problems? What are the statistical consequences of changes made for numerical reasons? How does one do inference on the output of nonstandard optimization problems? At the heart of these new estimation and inference questions lies the need to design and understand estimators as the product of algorithms and optimization problems, not only the minimand of objective functions.

Research at the intersection of econometrics, machine learning and optimization requires a specific set of skills. This course aims at providing such skills, with a particular focus on the optimization underlying modern estimation problems and the interplay of optimization and inference. The objective of this course is not to survey all new machine learning methods, but to trivialize them by forming a deep and intuitive understanding of the underlying concepts and techniques.

This course develops a precise and intuitive understanding of the optimization and algorithmic methods underpinning the evaluation of mainstream estimators in econometrics and machine learning. The advantage of developing such a computational and optimization-conscious understanding of econometric methods is two-fold: first, approaching statistical objects as the output of an optimization problem or algorithm affords the econometrician a new and conceptually rich angle from which to study estimators; second, when considering modifications or extensions of mainstream estimators, the econometrician may develop efficient and modern algorithms for her or his new methods.

The course both covers basic results in optimization –it should be considered as self-contained from an optimization perspective– and aims at getting students to tackle questions on the research frontier of econometrics. When studying the background optimization material, we use well understood econometric methods as running examples so to efficiently present the optimization methods within a familiar econometric environment.

Students completing this class should expect to have formed a firm understanding of Monte Carlo methods, linear programming, integer programming, numerical linear algebra, and how these interface with econometric and statistical methods such as quantile regression, least squares regression, best subset selection, estimation of partially identified models, etc.

The objective is to foster new and fresh research questions by connecting the econometric and optimization concepts.

Quantile regression holds a special place as a running example throughout a large part of the course. One of the objectives of the class is to develop a deep understanding of this regression tool.

Prerequisites for this course are econometrics or statistics at the first-year graduate level. Students with a strong background in linear algebra and econometrics may take the course with permission of the instructor. No optimization background is assumed. This course is the first installment of the second year PhD sequence in the field of econometrics for PhD students in the Department of Economics at Harvard University. Students wishing to qualify for specialization in the field of econometrics must take the course for a letter grade. All students attending the class are strongly encouraged to take the course for a letter grade.

Grading

Students will have a research paper as a final project. The final grade is broken down as follows:

- Participation: 20%
- Exercises and Problem sets: 40%
- Final project: 40%

Lecture notes will be released before lectures. **Students are required to read the lecture notes before lecture** so to help with participation. Students should have settled on a final project topic before the end of week 8.

Sections are an integral part of the course. Students should make sure to attend sections.

Topics Covered

Each lecture note contains material covered during one or more classes, and will be made available before class on the course website. Students are required to read the lecture notes ahead of class.

Part 1

Lecture 1. Monte Carlo Methods

> This section introduces Monte Carlo sampling methods from scratch. Precise topics covered will include Metropolis-Hastings as a unifying method, simulated annealing, auxiliary variables, pseudo-marginal MCMC and approximate Bayesian computations. The basic theory of Markov chains is covered. Some elementary results in the theory of mixing times for MCMC samplers are covered.

Part 2

Lecture 1. Quantile Regression and Linear Programming

> Properties of the quantile regression estimator are explored from the optimization and geometric angle of linear programming. Conversely, linear programming is studied via the example of quantile regression. Core linear programming concepts, such as duality, and solution algorithms, such as the simplex method, are developed in the statistical language and flavor of quantile regression.

Lecture 2. Nonparametric Robust Inference and Duality

> Rank tests may be generalized to the regression set-up, where an immediate connection arises with the dual problem of quantile regression. This allows one to build tests and confidence intervals for quantile regression coefficients without having to estimate an unwieldy asymptotic covariance matrix or running the bootstrap. Exact regression rankscore tests, as well as their different asymptotic approximations, are studied and critiqued. Computational methods are investigated.

Lecture 3. <u>Optimal Transport and Network Flow Problems, with Applications to Latent Variables Models</u>

> Optimal transport problems have gathered a lot of traction in economics as they have found powerful applications both in economic theory and in econometrics. They are members of a special family of linear programs, called network flow problems. These problems have beautiful properties and accommodate primal-dual algorithms, which we investigate. Applications studied include sampling methods for indirect inference estimators, vector quantile regression, and latent variables models. We consider uses of optimal transport as a reduced form regression tool for empirical economic analysis. We investigate connections between the optimal transport and quantile regression problems.

Lecture 4. <u>Entropy Regularization and Modern Optimization Algorithms for Optimal Transport</u>

> Entropy regularization of the optimal transport problem has allowed it to be used in much larger, big data applications. We study iterative projections and Sinkhorn decomposition approaches to the computation of entropy regularized optimal transport, as well as the deep connections between the two seemingly unrelated approaches. The excellent statistical performance of the regularized estimator brought up the question of its statistical interpretation, and recent research has produced many interesting answers. For instance, one may interpret the entropic optimal transport as maximum likelihood deconvolution. We will revisit recent econometric applications of optimal transport in latent variables models under this new light.

Lecture 5. <u>Certifiably Optimal Econometric Methods, Instrumental Variables Quantile Regression and Integer Programming</u>

> Many popular "sparse" methods in econometrics and statistics, such as the famous lasso, are relaxations of combinatorially difficult problems. Originally, the relaxations were put forth because the exact, combinatorially difficult versions were considered intractable. We will explore the feasibility, using modern optimization methods and computational power, of the exact form of sparse estimators such as best subset selection and sparse PCA. We will cover instrumental variables quantile regression (IVQR), where mixed integer linear programming methods recover the exact value of the regression coefficient estimate of this non-smooth and non-convex optimization problem and discuss how this solution method relates to regression rankscore inference for IVQR.

Part 3

Lecture 1. Least Squares Regression and Numerical Linear Algebra

> We study the machinery and matrix analysis behind OLS. This allows us to better investigate identification under rank deficiency.

Lecture 2. Data Sketching and the Johnson-Lindenstrauss Lemma

> We give a brief overview of data sketching. When datasets are very large, the complexity of standard linear algebraic operations may be reduced at very little cost in numerical accuracy by the use of random projection. This brings up interesting questions for optimization-conscious inference.

Part 4

Lecture 1. Degrees of Freedom of Constrained Estimators

> We revisit the literature on the degrees of freedom of the lasso, and extend the results to the synthetic control method. In order to achieve this, we present general Lagrange multiplier theory for two-level optimization programs.

Lecture 2. Convex Duality and Robustness

> We explore the duality relationship between distributionally robust optimization problems on the one hand, and corresponding machine learning and econometric methods on the other.

Part 5

Lecture 1. Numerical Integration and Uniform Inference

> We recast the problem of obtaining rectangular confidence regions for subvectors as a multivariate numerical integration problem. Applications include uniformly valid inference for quantile regression, as well as weak-identification robust inference for instrumental variables ordinary linear regression and instrumental variables quantile regression.

Timeline

Lectures are available on the course website. Section numbers refer to lectures available on the course website. For instance, 0.1-0.2 corresponds to sections 1 and 2 of Lecture 0. Required readings must be done **before the associated lecture**.

date	material	required readings
Thursday, September 3	Introduction to the course, and basic Markov Chain Monte Carlo Methods	0.0-0.1
Thursday, September 3	Problem Set 1 released	*
Tuesday, September 8	Advanced Markov Chain Monte Carlo Methods	0.3
Thursday, September 10	Elementary Markov Chain Monte Carlo Theory I	0.2
Thursday, September 10	Problem Set 1 due, Problem Set 2 Released	*
Tuesday, September 15	Elementary Markov Chain Monte Carlo Theory II	0.4
Thursday, September 17	Topics in Inference With Generative Models: ABC with $\epsilon = 0$?!	[1], [2], [3]
Thursday, September 17	Problem Set 2 due, Problem Set 3 Released	*
Tuesday, September 22	Topics in Inference With Generative Models: Alternative Notions of Distance Between True and Synthetic Data	[4], [5], [6]
Thursday, September 24	Modern Introduction to Quantile Regression	1.0-1.1
Thursday, September 24	Problem Set 3 due, Problem Set 4 released	*
Tuesday, September 29	Quantile Regression and Linear Programming I: Geometric Properties and Duality	1.2-1.3
Thursday, October 1	Quantile Regression and Linear Programming II: Estimation and the Simplex Method	1.4
Thursday, October 1	Problem Set 4 due, Problem Set 5 released	*
Tuesday, October 6	Rank Tests, Permutation Tests, and Asymptotic Robustness of Exact Tests	2.0-2.3
Thursday, October 8	Regression Rankscore Inference	2.4
Thursday, October 8	Problem Set 5 due, Problem Set 6 released	*
Tuesday, October 13	Suggestions for Final Projects	*
Thursday, October 15	Optimal Transport and Network Flow Problems	3.0-3.1
Thursday, October 15	Problem Set 6 due	*
Tuesday, October 20	Reduced-Form Applied Econometrics with Optimal Transport Methods	[7]
Thursday, October 22	Primal-Dual Algorithms and Warm Starting Network Flow Problems, Deconvolution by Simulation Using Optimal Transport Methods	3.1.3 [8], [9]
Thursday, October 22	Problem Set 7 released	*
Tuesday, October 27	Entropy Regularization and Modern Versions of Dykstra's Algorithm, Statistical Interpretation of Entropy Regularization and Connection with Maximum Likelihood	4
Thursday, October 29	Integer Programming, Branch-and-Bound Strategies, and examples in econometrics: exact sparse methods, sparse PCA, interval-censored data	TBA
Thursday, October 29	Problem Set 7 due, Problem Set 8 released	*
Tuesday, November 3	Instrumental Variables Quantile Regression	[10]
Thursday, November 5	Computing OLS. Identification of Regression Coefficients in Singular Ordinary Least-Squares Regression	5
Thursday, November 5	Problem set 8 due, Problem set 9 released	*
Tuesday, November 10	The Johnson-Lindenstrauss Lemma and Applications of Randomized Linear Algebra in Econometrics	TBA
Thursday, November 12	On the Degrees of Freedom of Constrained Estimators	TBA, [11]-[14]
Thursday, November 12	Problem set 9 due	*
Tuesday, November 17	Convex Duality and Robustness	TBA
Thursday, November 19	Empirical Likelihood Methods	[15]-[16]
Thursday, November 19	Problem Set 10 released	*
Tuesday, November 24	Volume Computations and Econometric Applications	TBA
Thursday, December 1	Numerical Integration	TBA
Thursday, December 1	Problem Set 10 due	*
Tuesday, December 3	Numerical Integration and Applications to Simultaneous Inference	TBA

[1] Brubaker, Marcus, Mathieu Salzmann, and Raquel Urtasun. "A family of MCMC methods on implicitly defined manifolds." In Artificial intelligence and statistics, pp. 161-172. 2012.

[2] Graham, Matthew M., Amos J. Storkey. "Asymptotically exact inference in differentiable generative models". https://arxiv.org/pdf/1605.07826.pdf

[3] Diaconis, Persi, and Bernd Sturmfels. "Algebraic algorithms for sampling from conditional distributions." The Annals of statistics 26, no. 1 (1998): 363-397.

[4] Gourieroux, Christian, Alain Monfort, and Eric Renault. "Indirect inference." Journal of applied econometrics 8, no. S1 (1993): S85-S118.

[5] Kaji, Tetsuya, Elena Manresa, and Guillaume Pouliot. "An Adversarial Approach to Structural Estimation." arXiv preprint arXiv:2007.06169 (2020).

[6] Bernton, Espen, Pierre E. Jacob, Mathieu Gerber, and Christian P. Robert. "Inference in generative models using the Wasserstein distance." arXiv preprint arXiv:1701.05146 1, no. 8 (2017): 9.

[7] Daljord, Øystein, Mandy Hu, Guillaume Pouliot, and Junji Xiao. "The Black Market for Beijing License Plates." Chicago Booth Research Paper 19-25 (2019).

[8] Arellano, Manuel, and Stéphane Bonhomme. "Recovering Latent Variables by Matching." arXiv preprint arXiv:1912.13081 (2019).

[9] Mallows, Colin. "Deconvolution by simulation." In Complex Datasets and Inverse Problems, pp. 1-11. Institute of Mathematical Statistics, 2007.

[10] Pouliot, Guillaume A. "Instrumental variables quantile regression with multivariate endogenous variable." Unpublished manuscript (2019).

[11] Pouliot Guillaume A., and Zhen Xie. "The degrees of freedom of the synthetic control method". Unpublished manuscript (2020).

[12] Zou, Hui, Trevor Hastie, and Robert Tibshirani. "On the "degrees of freedom" of the lasso." The Annals of Statistics 35, no. 5 (2007): 2173-2192.

[13] Tibshirani, Ryan J., and Jonathan Taylor. "Degrees of freedom in lasso problems." The Annals of Statistics 40, no. 2 (2012): 1198-1232.

[14] Tibshirani, Ryan, and L. Wasserman. "Stein's unbiased risk estimate." Course notes from "Statistical Machine Learning (2015): 1-12.

[15] Imbens, Guido W. "Generalized method of moments and empirical likelihood." Journal of Business & Economic Statistics 20, no. 4 (2002): 493-506.

[16] Parente, Paulo MDC, and Richard J. Smith. "Recent developments in empirical likelihood and related methods." Annu. Rev. Econ. 6, no. 1 (2014): 77-102.