Empirical Methods for Program Evaluation and Networks

CEMFI Summer School, August 24 to 28th, 2015

Course Description

This course begins with an overview of methods of **covariate adjustment**, as used, for example, to adjust for differences in baseline characteristics across treated and control units in an experimental or observational study. Our focus will be on methods which are semiparametrically efficient. The theory of semiparametric efficiency bounds will be used to provide guidance on estimator choice and other dimensions of research design. Methods of covariate adjustment for binary-valued treatment variables are widely-available. We will selectively review these methods but also study procedures appropriate for settings where the policy variable of interest is continuously-valued and/or time-varying.

We will next turn to the analysis of **matching and assignment problems**. We will consider two questions. First, how can we identify and efficiently estimate a match production function? For example, we might wish to understand how variation in characteristics of one parent, holding those of the other fixed, influences child outcomes. Second, how can we identify and estimate the parameters which determine who matches with whom? Answering this second question will require us to develop some basic theory of U-Statistics.

Next we will review methods for identifying and estimating **peer group effects**. Our analysis will begin with a careful study of the linear-in-means model and then turn to less parametric approaches. We will also consider **neighborhood effects**. Since sorting into neighborhoods is mediated by the market for housing, new issues and research design opportunities arise relative to the analysis of other types of social spillovers. Finally we will study some basic, and quite recent, models of **network formation.** Some of this material will require the use of the EM-Algorithm, which we will also review.

Course Logistics

Instructor: Bryan Graham, Department of Economics, University of California – Berkeley

Email: bgraham@econ.berkeley.edu

Time: 3:30PM to 7PM daily

Office Hours: I will announce the time and location of office hours on the first day of class. I look forward to talking with all of you!

Prerequisites: The equivalent of a first year Ph.D. level sequence in econometrics. Specifically an understanding of probability and statistical inference at the level of Casella and Berger (1990), linear regression analysis at the level of Goldberger (1991) and some exposure to non-linear models. I will also assume a basic knowledge of applied matrix algebra.

Textbook: Lecture will be based on the assigned readings as well as my own lecture notes (some of which will be made available to students prior to class). Readings preceded by a [r] in the course outline are "required" (i.e., should be read prior to class), while those preceded by a [b] are for "background" (i.e., may be useful for students interested in additional material). A useful general purpose reference is the textbook by Wooldridge (2010).

Computation: The bulk of class will be devoted to the formal development of the material, albeit with empirical illustrations as well as ample discussions of the various practicalities of implementation. However I do intend to reserve some class time for actual practice with computation. Computational examples will be done in a mix of Stata, MATLAB and Python. Python is free. I recommend the Anaconda installation available for download at http://continuum.io/downloads. Some basic tutorials on installing and using Python, with a focus on economics applications, can be found online at http://quant-econ.net.

Course Outline

Date	Topic	Readings
M 8/24	**Covariate adjustment**	[b] Imbens & Wooldridge (2009)
	Binary treatments	[r] Graham, Pinto and Egel (2012, forthcoming)
		[b] Graham (2011a)
	Continuously-valued treatments	[b] Newey (1990), [r] Robins, Mark & Newey (1992)
		[b] Hirano & Imbens (2004), [b] Wooldridge (2004)
	Dynamic treatments	[b] Robins, Hernán & Brumback (2000)
		[r] Hernán, Brumback & Robins (2001)
Tu 8/25	**Matching Problems**	[r] Graham (2011b)
	Assignment problems	[r] Graham, Imbens & Ridder (2015)
	U-Statistics	[b] van der Vaart (1998, Chapters 11 & 12)
		[r] Ferguson (2006)
	U-Process Minimizers	[b] Honore & Powell (1994), [b] Graham (2013)
W 8/26	**Peer Effects**	[r] Manski (1993), [b] Angrist (2015)
	Quasi-experiments	[r] Angrist & Lang (2004)
	Network structure	[r] Bramoullé, Djebbari & Fortin (2009)
	Covariance restrictions	[r] Graham (2008)
	Reallocations with spillovers	[r] Graham, Imbens & Ridder (2010)
Th 8/27	**Neighborhood Effects**	[b] Benabou (1996), [r] Graham (In Preparation)
	Cross-city research designs	[r] Card and Rothstein (2007)
	Covariate adjustment designs	[b] Sharkey & Elwert (2011)
		[r] Wodtke, Harding & Elwert (2011)
	MTO Experiment	[b] Chetty, Hendren & Katz (2015)
F 8/28	**Network formation**	[b] Goldenberg, Zheng, Fienberg & Airoldi (2009)
		[r] Graham (2015)
	β-Model, Testing	[b] Blitzstein and Diaconis (2011)
		[b] Chatterjee, Diaconis and Sly (2011)
	Stochastic Block Models	[b] Snijders & Nowicki (1997)
		[r] Daudin, Picard & Robin (2008)
		[b] Gupta and Chen (2011)
	Dyadic link formation	[r] Graham (2014)

References

[1] Angrist, Joshua. (2015). "The perils of peer effects," *Labour Economics* 30: 98 – 108.

[2] Angrist, Joshua D. and Kevin Lang. (1994). "Does school integration generate peer effects? Evidence from Boston's Metco program," *American Economic Review* 94 (5): 1613 - 1634.

[3] Benabou, Roland. (1996). "Equity and efficiency in human capital investment: the local connection," *Review of Economic Studies* 63 (2): 237 - 264.

[4] Blitzstein, Joseph and Persi Diaconis. (2011). "A sequential importance sampling algorithm for generating random graphs with prescribed degrees," *Internet Mathematics* 6 (4): 489 - 522.

[5] Bramoullé, Yann, Habiba Djebbari, Bernard Fortin. (2009). "Identification of peer effects through social networks," *Journal of Econometrics* 150 (1): 41 – 55.

[6] Card, David and Jesse Rothstein. (2007). "Racial segregation and the black-white test score gap," *Journal of Public Economics* 91 (11-12): 2158 - 2184.

[7] Casella, George and Roger L. Berger. (1990). *Statistical Inference.* Belmont, CA: Duxbury Press.

[8] Chatterjee, Sourav, Persi Diaconis and Allan Sly. (2011). "Random graphs with a given degree sequence," *Annals of Applied Probability* 21 (4): 1400 - 1435.

[9] Chetty, Raj. Nathaniel Hendren and Lawrence F. Katz. (2015). "The effects of exposure to better neighborhoods on children: new evidence from the Moving to Opportunity experiment," *NBER Working Paper No. 21156.*

[10] Daudin, J-J., F. Picard and S. Robin. (2008). "A mixture model for random graphs," *Statistics and Computing* 18 (2): 193 - 183.

[11] Ferguson, Thomas S. (2006). "U-Statistics," Lecture Note for Statistics 200C, UCLA.

[12] Goldberger, Arthur S. (1991). *A Course in Econometrics.* Cambridge, MA: Harvard University Press.

[13] Goldenberg, Anna, Alice X. Zheng, Stephen E. Fienberg and Edoardo M. Airoldi. (2009). "A survey of statistical network models," *Foundations and Trends in Machine Learning* 2 (2): 129 - 233.

[14] Graham, Bryan S. (2008). "Identifying social interactions through conditional variance restrictions," *Econometrica* 76 (3): 643 - 660.

[15] Graham, Bryan S. (2011a). "Econometric methods for the analysis of assignment problems in the presence of complementarity and social spillovers," *Handbook of Social Economics* 1B: 965 - 1052 (J. Benhabib, A. Bisin, & M. Jackson, Eds.). Amsterdam: North-Holland.

[16] Graham, Bryan S. (2011b). "Efficiency bounds for missing data models with semiparametric restrictions," *Econometrica* 79 (2): 437 - 452.

[17] Graham, Bryan S. (2013). "Comparative static and computational methods for an empirical one-to-one transferable utility matching model," *Advances in Econometrics* 31: 153 - 181.

[18] Graham, Bryan S. (2014). "An empirical model of network formation: detecting homophily when agents are heterogenous," *NBER Working Paper No. 20341.*

[19] Graham, Bryan S. (2015). "Methods of Identification in Social Networks," *Annual Review of Economics* 7.

[20] Graham, Bryan S. (In Preparation). "Identifying and estimating neighborhood effects: a review," Manuscript being prepared for the *Journal of Economic Literature*. Latest draft will be distributed in class.

[21] Graham, Bryan S., Guido W. Imbens, and Geert Ridder (2010). "Measuring the effects of segregation in the presence of social spillovers: a nonparametric approach," *NBER Working Paper No. 16499.*

[22] Graham, Bryan S., Guido W. Imbens, and Geert Ridder (2015). "Complementarity and aggregate implications of assortative matching: a nonparametric analysis," *Quantitative Economics* 5 (1): 29 - 66.

[23] Graham, Bryan S., Cristine Campos de Xavier Pinto. (2012). "Inverse probability tilting for moment condition models with missing data," *Review of Economic Studies* 79 (3): 1053 - 1079.

[24] Graham, Bryan S., Cristine Campos de Xavier Pinto. (forthcoming). "Efficient estimation of data combination models by the method of auxiliary-to-study tilting (AST)," *Journal of Business & Economic Statistics.*

[25] Gupta, Maya R. and Yihua Chen. (2011). "Theory and Use of the EM Algorithm," *Foundations and Trends in Signal Processing* 4 (3): 223 - 296.

[26] Hernán, Miguel Angel, Babette Brumback, James M. Robins. (2001). "Marginal structural models to estimate the joint causal effect of nonrandomized treatments," *Journal of the American Statistical Association* 96 (454): 440 - 448.

[27] Hirano, Keisuke and Guido W. Imbens. (2004). "The propensity score with continuous treatments," *Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives*: 73 - 84 (A. Gelman & X. Meng, Eds.). New York: John Wiley & Sons.

[28] Bo E. Honoré, James L. Powell. (1994). "Pairwise difference estimators of censored and truncated regression models," *Journal of Econometrics* 64 (1): 241 - 278.

[29] Imbens, Guido W., and Jeffrey M. Wooldridge. (2009). "Recent developments in the econometrics of program evaluation," *Journal of Economic Literature* 47 (1): 5-86.

[30] Manski, Charles F. (1993). "Identification of endogenous social effects: the reflection problem," *Review of Economic Studies* 60 (3): 531 - 542.

[31] Newey, Whitney K. (1990). "Semiparametric efficiency bounds," *Journal of Applied Econometrics* 5 (2): 99 - 135.

[32] Robins, James M. Miguel Angel Hernán and Babette Brumback. (2000). "Marginal structural models and causal inference in epidemiology," *Epidemiology* 11 (5): 550 - 560.

[33] Robins, James M., Steven D. Mark and Whitney K. Newey. (1992). "Estimating exposure effects by modeling the expectation of exposure conditional on confounders," *Biometrics* 48 (2): 479 - 495.

[34] Sharkey, Patrick and Felix Elwert. (2011). "The legacy of disadvantage: multigenerational neighborhood effects on cognitive ability," *American Journal of Sociology* 116 (6): 1934 - 81.

[35] Snijders, Tom A. B. and Krzysztof Nowicki. (1997). "Estimation and prediction for stochastic blockmodels for graphs with latent block structures," *Journal of Classification* 14 (1): 75 - 100.

[36] van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge: Cambridge University Press.

[37] Wodtke, Geoffrey T., David J. Harding and Felix Elwert. (2011). "Neighborhood effects in temporal perspective: the impact of long-term exposure to concentrated disadvantage on high school graduation," *American Sociological Review* 76 (5): 713 - 736.

[38] Wooldridge, Jeffrey M. (2004). "Estimating average partial effects under conditional moment independence assumptions, *CEMMAP Working Paper CWP03/04.*

[39] Wooldridge, Jeffrey M. (2010). *Econometric Analysis of Cross Section and Panel Data.* Cambridge, MA: The MIT Press.