Syllabus

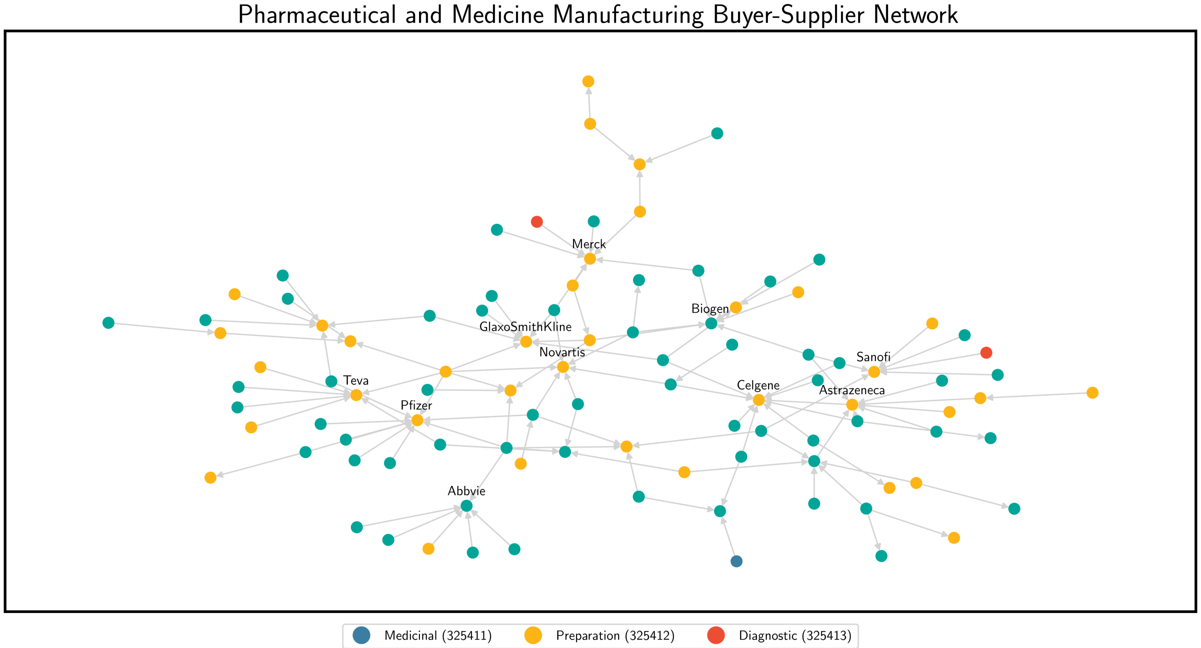
Ec 2147 Advanced Topics in Econometrics

Spring 2021

Course Description

The course begins with a review of the generalized method of moments (GMM) approach to estimation. Extensions of GMM to accommodation auxiliary moments, simulation and missing data are developed. The latter extension allows for the introduction of several canonical approaches to causal inference (e.g., covariate adjustment, instrumental variables). The theory of semiparametric efficiency bounds is also developed and applied to conditional moment problems. Next both linear and nonlinear panel data models are considered in some detail. The course ends with an introduction to some basic methods in the econometrics of games and networks. For each topic, theoretical development, computation and empirical application are all given due attention.

Pharmaceutical and Medicine Manufacturing Buyer-Supplier Network



Instructor: Bryan Graham, email: `bgraham@econ.berkeley.edu`

Time and Location: Tuesdays, 12PM to 2:45PM (Online/Zoom)

Office Hours: Thursday 5:00 PM - 7:00 PM (by Appointment) Sign up here.

Prerequisites: First year Ph.D.-level sequence in econometrics; linear algebra, multivariate calculus, basic probability and inference theory.

Course Webpage: Various instructional resources, including occasional lecture notes and Jupyter Notebooks, can be found on GitHub in the following repository https://github.com/bryangraham/

There is also a Harvard canvas site here. Some course materials made also be made available on this site.

Textbook: There is no mandatory text. Material will be delivered primarily through lecture and assigned papers. Good note taking is essential for successful performance in the class. Nevertheless I do recommend the following book as useful supplement to the material presented in lecture.

1. Wooldridge, Jeffrey M. (2011). *Econometric Analysis of Cross Section and Panel Data, 2^{nd} Ed*. Cambridge, MA: The MIT Press.

This is also a useful long term reference for anyone who anticipates undertaking empirical research. Also highly recommended is the *Handbook of Econometrics* chapter by Newey & McFadden (1994).

Additional books which you may find helpful include:

1. Arellano, Manuel. (2003). *Panel Data Econometrics.* Oxford: Oxford University Press.
2. van der Vaart, A. W. (1998). *Asymptotic Statistics.* Cambridge: Cambridge University Press.
3. Amemiya, Takeshi. (1985). *Advanced Econometrics.* Cambridge, MA: Harvard University Press.

Arellano (2003b) is an elegant and compact introduction to linear panel data analysis, which nicely complements Wooldridge (2010). van der Vaart (1998) is a standard introduction to large sample analysis. Amemiya (1985) is a somewhat dated, but nevertheless highly useful introduction to econometrics at an advanced level.

Grading: Grades will reflect performance on three separate homework assignments (75%) and regular classroom participation (25%). Two of the homeworks will be problems sets with a combination of pencil and paper and computation work. The final assignment is a short paper of 10 to 15 pages. You are free to work in groups on assignments (indeed encouraged!) but each student must submit an individual write-up, accompanying computer code, and executed output. Specifically you are expected to write your own solutions and your own code. Consult with classmates about the structure of solutions, keys steps and ideas, algorithms, program structure etc. Do not "cut-and-paste" joint work. Problem sets are graded both on their correctness and their clarity.

Computation: All computational work should be completed in Python. Python is a widely used general purpose programming language with good functionality for scientific computing. I recommend the Anaconda distribution, which is available for download at http://continuum.io/downloads. Some basic tutorials on installing and using Python, with a focus on economic applications, can be found online at `https://python.quantecon.org/`.

Good books for learning Python, with some coverage of statistical applications, are Guttag (2013), VanderPlas (2017) and McKinney (2017). The first book is an excellent introduction to computer science as well as Python, the last book is heavily focused on the pandas module.

Any code I will provide will execute properly on Python 3.6. Students wishing to work with another technical computing environment (e.g., MATLAB, Julia, Fortran 95, C++, R, etc.) should speak with me. This will be allowed at my discretion. There are a large number of useful resources available for learning Python.

While issues of computation may arise from time to time during lecture, I will not teach Python programming. *This is something you will need to learn outside of class.* I do not expect this to be easy. I ask that those students with strong backgrounds in technical computing assist classmates with less experience.

Course Goals, Teaching Philosophy and Expectations: Second year graduate coursework marks a transition from highly structured learning to more individualized learning and research. While I hope that by the end of this course you will have mastered a few additional tools that you can directly apply to your own research, this is not a bag of tricks or recipe type course. My view is that it is more valuable to work through a small number of examples in some detail, than to superficially cover many topics. Real world empirical research rarely fits into a neat template for which I can teach you the correct tool. Even if your main focus is on empirical work, some methodological innovation/adaptation will be required of you to write interesting papers. This course is more about how to go about formulating solutions to econometric problems, than it is about surveying extant solutions to econometric problems. For students who are considering research in theoretical econometrics, I hope to give you a flavor of how to find interesting ideas and some pointers on open questions. While at times we will delve into technical details, the development of ideas will be more informal: calculations/derivations/arguments versus assumptions/theorems/proofs. Of course development of ideas in terms of assumptions/theorems/proofs is required for publishing econometrics research. There is a gap. I can provide pointers on the required self-study needed to rigorously fill out the more skeletal arguments we will develop in class. Finally, don't be discouraged if you don't understand everything. Research involves intense thinking precisely about ideas which you do not understand. Confusion is baked into the process of academic research; acquire a taste for it and have fun.

Additional notes: I prefer to avoid having substantive communications by e-mail. Please limit e-mail use to short yes/no queries. I am unlikely to read or respond to a long/complex e-mail. Do feel free to chat with me immediately before class (I will aim to join the Zoom room a bit early and linger a bit late). For longer questions please make use of my office hours. This is time specifically allocated for your use; please come by! I look forward to getting to know all of you.

Course Outline

DATE	TOPIC	READINGS/NOTES t = theory, a = application, b = background
Tu 1/26	GMM: Sequential Estimation	[t] Newey (1984) [b] Wooldridge (2010, Ch. 14)
Tu 2/2	GMM: Auxiliary Moments	[t] Back & Brown (1993), [t] Imbens & Lancaster (1994) [b] Hailong & Schmidt (1999)
Tu 2/16	GMM: Semiparametric Efficiency Bounds	[t] Newey (1990), [t] Chamberlain (1987) [b] Tsiatis (2006, Chapters 2 to 4) [b] Brown & Newey (1998)
Tu 2/9	GMM: Missing Data	[a] Graham et al. (2012), [t] Graham (2011) [b] Angrist et al. (1996), [a] Abadie (2003) [b] Prokhorov & Schmidt (2009)
Tu 2/23	GMM: Simulation	[a] Laffont et al. (1995), [a] Keane (2015) [b] Pakes & Pollard (1991)
Tu 3/2	EM Algorithm	[t] Gupta & Chen (2010), [b] Murphy (2012, Chapter 11) [a] Imbens & Rubin (1997), [a] Porter (1983) [a] Hahn & Moon (2010)
Tu 3/9	Linear Panel Data: Static Models	[t] Chamberlain (1984), [a] Griliches & Mairesse (1998) [b] Wooldridge (2010, Ch. 10) [a] Rosenzweig & Wolpin (1986), [a] Jakubson (1991)
Tu 3/16	*Wellness Day* (No Class)	
Tu 3/23	Linear Panel Data: Dynamic Models	[t] Arellano (2003b, Chapters 7 & 8), [b] Wooldridge (2010, Ch. 11) [a] Blundell & Bond (2000) [a] Olley & Pakes (1996), [a] de Loecker & Warzynski (2012)
Tu 3/30	Supermodular games	[t] Uetake & Watanabe (2013) , [a] Jia (2008) [a] Miyauchi (2016)
Tu 4/6	Nonlinear Panel Data: Static Models	[t] Arellano (2003a), [t] Bonhomme (2012) [b] Chamberlain (1980), [b] Manski (1987) [b]
Tu 4/13	Nonlinear Panel Data: Dynamic Models	[t] Chamberlain (1985), [t] Honoré & Kyriazidou (2000) [t] Honoré & Weidner (2020) [a] Card & Sullivan (1988), [a] Card & Hyslop (2005)
Tu 4/20	Networks: Dyadic Regression	[t] Graham (2020b, Sections 4 - 6) [t] Graham (2020a) [b] van der Vaart (1998, Chapters 11 & 12)
Tu 4/27	Networks: Simulation/Testing	[t] Blitzstein & Diaconis (2011) [t] Pelican & Graham (2020) [b] Graham & Pelican (2020)

References

Abadie, A. (2003). Semiparametric instrumental variables estimation of treatment response models. *Journal of Econometrics*, 113(2), 231 – 263.

Amemiya, T. (1985). *Advanced Econometrics*. Harvard University Press.

Angrist, J. D., Imbens, G. W., & Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, 91(434), 444 – 455.

Arellano, M. (2003a). Discrete choices with panel data. *Investigaciones Económicas*, 27(3), 423 – 458.

Arellano, M. (2003b). *Panel Data Econometrics*. Oxford: Oxford University Press.

Back, K. & Brown, D. P. (1993). Implied probabilities in gmm estimators. *Econometrica*, 61(4), 971 – 75.

Blitzstein, J. & Diaconis, P. (2011). A sequential importance sampling algorithm for generating random graphs with prescribed degrees. *Internet Mathematics*, 6(4), 489 – 522.

Blundell, R. & Bond, S. (2000). Gmm estimation with persistent panel data: an application to production functions. *Econometric Reviews*, 19(3), 321 – 340.

Bonhomme, S. (2012). Functional differencing. *Econometrica*, 80(4), 1337 – 1385.

Brown, B. W. & Newey, W. K. (1998). Efficient semiparametric estimation of expectations. *Econometrica*, 66(2), 453 – 464.

Card, D. & Hyslop, D. R. (2005). Estimating the effect of a time-limited subsidy for welfare-leavers. *Econometrica*, 73(6), 1723 – 1770.

Card, D. & Sullivan, D. (1988). Measuring the effect of a subsidized training program on movements in and out of employment. *Econometrica*, 56(3), 497 – 530.

Chamberlain, G. (1980). Analysis of covariance with qualitative data. *Review of Economic Studies*, 47(1), 225 – 238.

Chamberlain, G. (1984). *Handbook of Econometrics*, volume 2, chapter Panel Data, (pp. 1247 – 1318). North-Holland: Amsterdam.

Chamberlain, G. (1985). *Longitudinal Analysis of Labor Market Data*, chapter Heterogeneity, omitted variable bias, and duration dependence, (pp. 3 – 38). Cambridge University Press: Cambridge.

Chamberlain, G. (1987). Asymptotic efficiency in estimation with conditional moment restrictions. *Journal of Econometrics*, 34(3), 305 – 334.

de Loecker, J. & Warzynski, F. (2012). Markups and firm-level export status. *American Economic Review*, 102(6), 2437 – 2471.

Graham, B. S. (2011). Efficiency bounds for missing data models with semiparametric restrictions. *Econometrica*, 79(2), 437 – 452.

Graham, B. S. (2020a). *The Econometrics of Social and Economic Networks*, chapter Dyadic regression, (pp. 25– 40). Elsevier: Amsterdam.

Graham, B. S. (2020b). *Handbook of Econometrics*, volume 7, chapter The econometric analysis of networks, (pp. 111 – 218). North-Holland: Amsterdam, 1st edition.

Graham, B. S. & Pelican, A. (2020). *The Econometrics of Social and Economic Networks*, chapter Testing for externalities in network formation using simulation",, (pp. 63 – 82). Elsevier: Amsterdam.

Graham, B. S., Pinto, C., & Egel, D. (2012). Inverse probability tilting for moment condition models with missing data. *Review of Economic Studies*, 79(3), 1053 – 1079.

Griliches, Z. & Mairesse, J. (1998). *Econometrics and Economic Theory in the 20th Century*, chapter Production functions: the search for identification, (pp. 169 – 203). Cambridge University Press: Cambridge.

Gupta, M. R. & Chen, Y. (2010). Theory and use of the em algorithm. *Foundations and Trends in Signal Processing*, 4(3), 223 – 296.

Guttag, J. V. (2013). *Introduction to Computation and Programming Using Python*. Cambridge, MA: The MIT Press.

Hahn, J. & Moon, H. R. (2010). Panel data models with finite number of multiple equilibria. *Econometric Theory 26*, 26(3), 863 – 881.

Hailong, Q. & Schmidt, P. (1999). Improved instrumental variables and generalized method of moments estimators. *Journal of Econometrics*, 91(1), 145 – 169.

Honoré, B. E. & Kyriazidou, E. (2000). Panel data discrete choice models with lagged dependent variables. *Econometrica*, 68(4), 839 – 874.

Honoré, B. E. & Weidner, M. (2020). *Moment conditions for dynamic panel logit models with fixed effects*. Technical Report arXiv:2005.05942, arXiv.

Imbens, G. W. & Lancaster, T. (1994). Combining micro and macro data in microeconometric models. *Review of Economic Studies*, 61(4), 665 – 680.

Imbens, G. W. & Rubin, D. B. (1997). Estimating outcome distributions for compliers in instrumental variable models. *Review of Economic Studies*, 64(4), 555 – 574.

Jakubson, G. (1991). Estimation and testing of the union wage effect using panel data. *Review of Economic Studies*, 58(5), 971 – 991.

Jia, P. (2008). What happens when wal-mart comes to town: an empirical analysis of the discount retailing industry. *Econometrica*, 76(6), 1263 – 1316.

Keane, M. P. (2015). *The Oxford Handbook of Panel Data*, chapter Panel discrete choice models of consumer demand, (pp. 551 – 582). Oxford University Press.

Laffont, J.-J., Ossard, H., & Vuong, Q. (1995). Econometrics of first-price auctions. *Econometrica*, 63(4), 953 – 980.

Manski, C. F. (1987). Semiparametric analysis of random effects linear models from binary panel data. *Econometrica*, 55(2), 357 – 362.

McKinney, W. (2017). *Python for Data Analysis*. Cambridge: O'Reilly.

Miyauchi, Y. (2016). Structural estimation of a pairwise stable network with nonnegative externality. *Journal of Econometrics*, 195(2), 224 – 235.

Murphy, K. P. (2012). *Machine Learning: A Probabilistic Perspective*. Cambridge, MA: The MIT Press.

Newey, W. K. (1984). A method of moments interpretation of sequential estimators. *Economics Letters*, 14(2-3), 201 – 206.

Newey, W. K. (1990). Efficient instrumental variables estimation of nonlinear models. *Econometrica 58*, 58(4), 809 – 837.

Newey, W. K. & McFadden, D. (1994). *Handbook of Econometrics*, volume 4, chapter Large sample estimation and hypothesis testing, (pp. 2111 – 2245). North-Holland: Amsterdam.

Olley, G. S. & Pakes, A. (1996). The dynamics of productivity in the telecommunications equipment industry. *Econometrica*, 64(6), 1263 – 1297.

Pakes, A. & Pollard, D. (1991). Simulation and the asymptotics of optimization estimators. *Econometrica*, 57(5), 1027 – 1057.

Pelican, A. & Graham, B. S. (2020). *An optimal test for strategic interaction in social and economic network formation between heterogeneous agents*. Working paper, University of California - Berkeley.

Porter, R. H. (1983). A study of cartel stability: the joint executive committee, 1880-1886. *Bell Journal of Economics*, 14(2), 301 – 314.

Prokhorov, A. & Schmidt, P. (2009). Gmm redundancy results for general missing data problems. *Journal of Econometrics*, 151(1), 47 – 55.

Rosenzweig, M. R. & Wolpin, K. I. (1986). Evaluating the effects of optimally distributed public programs: child health and family planning interventions. *American Economic Review, vol. 76, no. 3, 1986, pp. 470–482*, 76(3), 470 – 482.

Tsiatis, A. A. (2006). *Semiparametric Theory and Missing Data*. New York: Springer.

Uetake, K. & Watanabe, Y. (2013). Estimating supermodular games using rationalizable strategies. *Advances in Econometrics*, 31(1), 233 – 247.

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

VanderPlas, J. (2017). *Python Data Science Handbook*. Boston: O'Reilly.

Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: MIT Press, 2nd edition.