Quantitative Empirical Methods in Information Systems

Professor: Dr. Qiang (David) Gao
Email: Qiang.Gao@baruch.cuny.edu
Class Meetings: Wed 1:00-3:00 PM
Class Room: TBD
Department of Computer Information Systems
Office: VC-11-244
Office Hours: Wed 12:30-1 PM by appointment

1. Overview:

It is increasingly common to see research studies in the field of Information Systems utilizing methods from economics, computational linguistics, and computer sciences. Also, the recent trend in literature shows the applications with advanced empirical methods, such as discrete choice models, Bayesian models, latent variable and hidden Markov models, and predictive models. The objective of this doctoral level course is to introduce these methods and their applications to the Ph.D. students in the IS and other majors. The class will survey the existing literature for typical applications of modeling approaches.

The class will meet once a week for around 2 hours. After taking the course, the students should be able to understand and implement learned empirical methods in producing an original research proposal.

2. Recommended Textbooks:

- Willian H. Greene (2012) Econometric Analysis (Seventh Edition). Pearson Education Limited. (International edition available on Amazon.com)
- Jeffrey M. Wooldridge. (2010) Econometric Analysis of Cross Section and Panel Data (Second Edition).
- Joshua D. Angrist, Jorn-Steffen Pischke (2009) Mostly Harmless Econometrics: An Empiricist's Companion. Princeton University Press.
- Daniel Jurafsky, J. H Martin (2023). Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition. Third Edition
- Maria L. Rizzo (2008) Statistical Computing with R. Chapman & Hall/CRC.

3. Course Structure and Expectations

The class includes lectures, readings, independent computational work, and presentations. Class attendance is critical as interaction with peers is an essential aspect of this course and the learning process associated with it.

- I will communicate with you through your Baruch College e-mail. I expect that you read it at least once a day.
- Preparation before each class: having read the required material carefully, prepared the presentation slides, and completed exercises if any.
- Classroom norms and etiquette: You are expected join each class on time and actively participate in discussions. We will often use computers for exercises or searches in class. Please keep other non-academic materials, including social media closed.

4. Class Assessment:

(1) Grading:

The grade is a combination of:

- *Weekly summary reports 10%*

Students should submit a brief report that summarizes the week's readings, starting from the second week. In the summary report, students should briefly describe each paper, and highlight 2-3 strengths and 1-2 shortcomings of the paper (how the authors can improve the paper). Any paper marked with a "*" is optional reading and is not required for a summary.

- *In-class presentation and discussion 20%*

In every class, each student is assigned to present one selected paper with a set of PPT slides. Each class presentation should take around 15 minutes and leaves 5 minutes for discussion. The instructor will assign papers to students' next presentation at end of each meeting, according to the students' preference. Any paper marked with a "#" is open for student presentation

- *Two homework sets 10%*

Two simple econometrics homework sets require the students to use two commonly used software, Stata and R, respectively.

- *Two computational problem sets 30%*

At least two complicated problems sets require the students to use R programming language to solve simulation-based models. The given papers are directly related to the computational problems.

- *Research Paper (Proposal) 30%*

Each student must produce a research-in-progress paper (research proposal) that has the quality to be accepted by major IS conferences, such as WISE and CIST. Around 7th week of this course, the students must formulate 3-4 research questions that can be answered by proposed empirical analyses. In the later weeks, the instructor will work with the student to select one of them and to develop it into a research paper. The student will present the research paper in the last week. The evaluation of the research paper depends on the potential impacts of the research topic, comprehensiveness and depth of the literature review, coherence, and logic of the proposed hypotheses, the correctness of the research methods/design, and strength of the preliminary results (if there is any):

  - Instructor evaluation: 80%
  - Peer evaluation: 20%

(2) Late and Assignment Incompletion Policy

Assignments are due prior to 12:59 p.m. on the dates announced in classes. Late submission of problem sets will lead to a two-point reduction for missing the deadline. two additional points for a one-week delay, and zero thereafter. Each incompletion of assigned reading and presentation will lead to a two-point reduction.

5. Academic Integrity:

The CIS Department fully supports Baruch College's policy on Academic Honesty, which states, in part:

"*Academic dishonesty is unacceptable and will not be tolerated. Cheating, forgery, plagiarism and collusion in dishonest acts undermine the college's educational mission and the students' personal and intellectual growth. Baruch students are expected to bear individual responsibility for their work, to learn the rules and definitions that underlie the practice of academic integrity, and to uphold its ideals. Ignorance of the rules is not an acceptable excuse for disobeying them. Any student who attempts to compromise or devalue the academic process will be sanctioned.* "

The use of AI (ChatGPT and similar) for coursework and assignments is strictly prohibited. This includes, but is not limited to, the use of AI-generated text, speech, programming code or images, as well as the use of AI tools or software to complete any portion of a project and assignment. Any use of AI tools to complete your work or a portion of your work will result in a grade of 0.

A report of suspected academic dishonesty will be sent to the Office of the Dean of Students. Additional information and definitions can be found at:
https://www.baruch.cuny.edu/academic/academic_honesty.html

6. Accommodation for Disability and Religious Observance

Baruch has a continuing commitment to providing reasonable accommodations for students with disabilities. Students with disabilities who may need some accommodation in order to fully participate in this class should contact Student Disability Services as soon as possible at disability.services@baruch.cuny.edu or call 646/312-4590.

At Baruch, we acknowledge that as a student, you are balancing many demands. During the semester, if you start to experience personal difficulties or stressors that are interfering with your academic performance or day to day functioning, please consider seeking free and confidential support at the Baruch College Counseling Center. For more information or to make an appointment, please visit their website at https://studentaffairs.baruch.cuny.edu/counseling/ or call 646-312-2155. If it's outside of business hours (Monday-Friday 9-5pm) and you need immediate assistance, please call 1-888-NYC-WELL (888-692-9355). If you are concerned about one of your classmates, please share that concern by filling out a Campus Intervention Team form at https://studentaffairs.baruch.cuny.edu/campus-intervention-team.

If you wish to be absent from class to observe a religious holiday, make arrangements in advance with the instructor.

7. Tentative Schedule and Topics

Dates	Topics	Reading	Homework
Week 1	IS Research Methods: Introduction and Examples	List 1	
Week2	IT Productivity and Spillover	List 2	
Week 3	E-Commerce	List 3	Problem set using Stata
Week 4	Social Media	List 4	
Week 5	Digital Goods	List 5	Problem set using R
Week 6	Platform Economy	List 6	
Week 7	Search Engine	List 7	Research proposal ideas
Week 8-9	Hidden Markov Models	List 8	Computational problem set: Hidden Markov model
Week 10-11	Spatial and Network Models	List 9	Computational problem set: Spatial econometric model
Week 12-13	Text Analysis, Big Data Analytics, and Predictive Models	List 10	
Week 14 (if 15 weeks schedule)	Economics of AI	List 11	
Week 14-15 (if 16 weeks schedule)	(Static) Discrete Choice Models	List 12	Computational problem set: BLP
Week 16	Research Paper Presentations		

Note: Reading list will be updated at the beginning of semester to reflect students' backgrounds.

List 1. IS Research Methods: Introduction and Examples

PE Todd (2006) Matching Estimators, University of Pennsylvania. Available online: http://athena.sas.upenn.edu/petra/papers/mpalgrave2.pdf.

EA Staurt (2010) Matching Methods for Causal Inference: A Review and A Look Forward. *Statistical Science* 25(1): 1-21.

Bertrand M, Duflo E, Mullainathan S (2004) How much should we trust difference-in-differences estimates? *Quarterly Journal of Economics* 119(1): 249-275.

DS Lee, T Lemieux (2010) Regression Discontinuity Designs in Economics. *Journal of Economic Literature* 48: 281-355.

#G Adomavicius, JC Bockstedt, SP Curley, J Zhang (2013) Do Recommender Systems Manipulate Consumer Preferences? A Study of Anchoring Effects. *Information Systems Research* 24(4): 956-975.

#W Khern-am-nuai, K Kannan, H Ghasemkhani (2018) Extrinsic versus Intrinsic Rewards for Contributing Reviews in an Online Platform. *Information Systems Research* 29(4): 871-892.

#M Sun (2012) How Does the Variance of Product Ratings Matter? *Management Science* 58(4): 696-707.

List 2: IT Productivity and Spillover

TF Bresnahan, E Brynjolfsson, LM Hitt (2002) Information Technology, Workplace Organization, and the Demand for Skilled Labor: Firm-Level Evidence. *Quarterly Journal of Economics* 117(1): 339-376.

#P Tambe, LM Hitt, E Brynjolfsson (2012) The Extroverted Firm: How External Information Practices Affect Innovation and Productivity. *Management Science* 58(5): 843-859.

#Z Cheng, BR Nault (2007) Industry Level Supplier-Driven IT Spillover. *Management Science* 53(8): 1199-1216.

#YB Chang, V Gurbaxani (2012) The Impact of IT-Related Spillovers on Long-Run Productivity: An Empirical Analysis. *Information Systems Research* 23(3): 868-886.

*P Tambe, LM Hitt (2014) Measuring Information Technology Spillovers. *Information Systems Research* 25(1): 53-71.

*Z Cheng, BR Nault (2012) Relative Industry Concentration and Customer-Driven IT Spillovers. *Information Systems Research* 23(2): 340-355.

*L Wu, F Jin, LM Hitt (2018) Are All Spillovers Created Equal? A Network Perspective on Information Technology Labor Movements. *Management Science* 64(7): 3168-3186.

List 3: E-Commerce

K Xu, J Chan, A Ghose, SP Han (2016) Battle of the Channels: The Impact of Tablets on Digital Commerce. *Management Science* 63(5): 1469-1492.

#TF Tan, S Netessine, L Hitt (2017) Is Tom Cruise Threatened? An Empirical Study of the Impact of Product Variety on Demand Concentration. *Information Systems Research* 28(3): 643-660.

#A Kumar, A Mehra, S Kumar (2019) Why Do Stores Drive Online Sales? Evidence of Underlying Mechanisms from a Multichannel Retailer. *Information Systems Research* 30(1): 319-338.

#DR Bell, S Gallino, A Moreno (2019) Offline Showrooms in Omnichannel Retail: Demand and Operational Benefits. *Management Science* 64(4): 1629-1651.

*U Narang, V Shankar (2019) Mobile App Introduction and Online and Offline Purchases and Product Returns. Marketing Science 38(5).

*E Overby, C Forman (2015) The Effect of Electronic Commerce on Geographic Purchasing Patterns and Price Dispersion. *Management Science* 61(2): 431-453.

*ET Anderson, NM Fong, DI Simester, CE Tucker (2010) How Sales Taxes Affect Customer and Firm Behavior: The Role of Search on the Internet. *Journal of Marketing Research* vol. XLVII (April 2010): 229-239.

*G Oestreicher-Singer, A Sundararajan (2012) Recommendation Networks and the Long Tail of Electronic Commerce. *MIS Quarterly* 36(1): 65-83.

List 4: Social Media

D Mayzlin, Y Dover, J Chevalier (2014) Promotional Reviews: An Empirical Investigation of Online Review Manipulation. *American Economic Review* 104(8): 2421-2455.

#H Chen, P De. YJ Hu (2015) IT-Enabled Broadcasting in Social Media: An Empirical Study of Artists' Activities and Music Sales. *Information Systems Research* 26(3): 513-531.

#Y-J Lee, K Hosanagar, Y Tan (2015) Do I Follow My Friends or the Crowd? Information Cascades in Online Movie Ratings. *Management Science* 61(9): 2241-2258.

#M Anderson, J Magruder (2012) Learning from the Crowd: Regression Discontinuity Estimates of the Effects of an Online review Database. *The Economic Journal* 122: 957-989.

*X Li (2018) Impact of Average Rating on Social Media Endorsement: The Moderating Role of Rating Dispersion and Discount Threshold. *Information Systems Research* 29(3).

*W Jabr, Z Zheng (2014) Know Yourself and Know Your Enemy: An Analysis of Firm Recommendations and Consumer Reviews in a Competitive Environment. *MIS Quarterly* 38(3): 635-654.

*M Luca, G Zervas (2017) Fake It till You Make It: Reputation, Competition, and Yelp Review Fraud. *Management Science* 62(12): 34312-3427.

*WW Moe DA Schweidel (2012) Online Product Opinions: Incidence, Evaluation, and Evolution. *Marketing Science* 31(3): 372-386.

List 5: Digital Goods

H Krijestorac, R Garg, V Mahajan (2020) Cross-Platform Spillover Effects in Consumption of Viral Content: A Quasi-Experimental Analysis Using Synthetic Controls. *Information Systems Research* 31(2): 449-472.

#R Belo, P Ferreira, MG de Matos, F Reis (2019) The Impact of Time Shifting on TV Consumption and Ad Viewership. *Management Science* 65(7): 3216-3234.

#B Danaher, S Dhanasobhon, MD Smith, R Telang (2010) Converting Pirates without Cannibalizing Purchasers: The Impact of Digital Distribution on Physical Sales and Internet Piracy. *Marketing Science* 29(6): 1138-1151.

#S Aral, D Walker (2011) Creating Social Contagion Through Viral Product Design: A Randomized Trial of Peer Influence in Networks. *Management Science* 57(9): 1623-1639.

*W Wen, M Ceccagnoli, C Forman (2015) Opening Up Intellectual Property Strategy: Implications for Open Source Software Entry by Start-up Firms. *Management Science* 62(9): 2668-2691.

List 6: Platform Economy

L Yang, Z Wang, J Hahn (2020) Scarcity Strategy in Crowdfunding: An Empirical Exploration of Reward Limits. *Information Systems Research*, forthcoming.

#H Chen, P De, Y Hu, B-H H (2014) Wisdom of Crowds: The Value of Stock Opinions Transmitted Through Social Media. *Review of Financial Studies* 27(5): 1367-1403.

#G Zervas, D Proserpio, JW Byers (2017) The Rise of the Sharing Economy: Estimating the Impact of Airbnb on the Hotel Industry. *Journal of Marketing Research* LIV: 687-705.

#MD Leung (2019) Learning to Hire? Hiring as a Dynamic Experiential Learning Process in on Online Market for Contract Labor. *Management Science* 64(12): 5651-5668.

*Z Wei, M Lin (2017) Market Mechanism in Online Peer-to-Peer Lending. *Management Science* 63(12).

*M Lin, S Viswanathan (2015) Home Bias in Online Investments: An Empirical Study of an Online Crowdfunding Market. *Management Science* 62(5): 1393-1414.

*J Zhang, P Liu (2012) Rational Herding in Microloan Markets. *Management Science* 58(5): 892-912.

Research paper: ideas

List 7: Search Engine

J Gong, V Abhishek, B Li (2017) Examining the Impact of Keyword Ambiguity on Search Advertising Performance: A Topic Model Approach. *MIS Quarterly* 42(3): 805-829.

#M Chesnes, W Dai, GZ Jin (2017) Banning Foreign Pharmacies from Sponsored Search: The Online Consumer Response. *Marketing Science* 36(6): 879-907.

#S Narayanan, K Kalyanam (2015) Position Effects in Search Advertising and their Moderators: A Regression Discontinuity Approach. *Marketing Science* 34(3): 388-407.

#A Ghose, S Yang (2009) An Empirical Analysis of Search Engine Advertising: Sponsored Search in Electronic Markets. *Management Science* 55(10): 1605-1622.

*A Agarwal, K Hosanagar, MD Smith (2011) Location, Location, Location: An Analysis of Profitability of Position in Online Advertising Markets. *Journal of Marketing Research* XLVIII (Dec 2011): 1057-1073.

*Y Chen, S Yao (2017) Sequential Search with Refinement: Model and Application with Click-Steam Data. *Management Science* 63(12): 4345-4365.

List 8. Hidden Markov Models (2 Weeks)

Y Huang, S Jasin, P Manchanda (2019) "Level Up": Leveraging Skill and Engagement to Maximize Player Game-Play in Online Video Games. *Information Systems Research* 30(3): 927-947.

#R Bapna, P Goes, KK Wei, Z Zhang (2011) A Finite Mixture Logit Model to Segment and Predict Electronic Payments System Adoption. *Information Systems Research* 22(1): 118-1333.

#L Yan, Y Tan (2014) Feeling Blue? Go Online: An Empirical Study of Social Support Among Patients. *Information Systems Research* 25(4): 690-709.

#W Chen, X Wei, KX Zhu (2018) Engaging Voluntary Contributions in Online Communities: A Hidden Markov Model. *MIS Quarterly* 42(1): 83-100.

*N Sahoo, PV Singh, T Mukhopadhyay (2012) A Hidden Markov Model for Collaborative Filtering. *MIS Quarterly* 36(4): 1329-1356.

*PV Singh, N Sahoo, T Mukhopadhyay (2014) How to Attract and Retain Readers in Enterprise Blogging? *Information Systems Research* 25(1): 35-42.

S Gupta, PK Chintagunta (1994) On Using Demographic Variables to Determine Segment Membership in Logit Mixture Models. *Journal of Marketing Research* XXXI (Feb 1994): 128-136.

PV Singh, Y Tan, and N Youn (2011): A Hidden Markov Model of Developer Learning Dynamics in Open Source Software Projects. *Information Systems Research* 22(4): 790-807.

O Netzer, JM Lattin, V Srinivasan (2008) A Hidden Markov Model of Customer Relationship Dynamics. *Marketing Science* 27(2): 185-204.

LR Rabiner (1989) A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. *Proceedings of the IEEE* 77(2): 257-286.

List 9. Spatial and Network Models (2 Weeks)

SK Shriver, HS Nair, R Hofstetter (2013) Social Ties and User-Generated Content: Evidence from an Online Social Network. *Management Science* 59(6): 1425-1443.

#Z Shi, GM Lee, AB Whinston (2016) Toward A Better Measure of Business Proximity: Topic Modeling for Industry Intelligence. *MIS Quarterly* 40(4): 1035-1056.

#K Zhang, S Bhattacharyya, S Ram (2016) Large-Scale Network Analysis for Online Social Brand Advertising. *MIS Quarterly* 40(4): 849-868.

#Y Lu, K Jerath, PV Singh (2013) The Emergence of Opinion Leaders in a Networked Online Community: A Dyadic Model with Time Dynamics and A Heuristic for Fast Estimation. *Management Science* 59(8): 1783-1799.

#MG de Matos, P Ferreira, D Krackhardt (2014) Peer Influence in the Diffusion of iPhone 3G over a Large Social Network. *MIS Quarterly* 38(4): 1103-1133.

*B Zhang, PA Pavlou, R Krishnan (2018) On Direct vs. Indirect Peer Influence in Large Social Networks. *Information Systems Research* 29(2): 292-314.

WR Hartmann, P Manchanda, H Nair, M Bothner, P Dodds, D Godes, K Hosanagar, C Tucker (2008)

Modeling Social Interactions: Identification, Empirical Methods and Policy Implications. Marketing Letters 19: 287-304.

S Yang, GM Allenby (2003) Modeling Interdependent Consumer Preferences. *Journal of Marketing Research* Vol. XL: 282-294.

TE Smith, JP LeSage (2004) A Bayesian Probit Model with Spatial Dependencies. *Spatio and Spatiotemporal Econometrics* 18: 127-160.

List 10. Text Analysis and Big Data Analytics

GM Lee, S He, J Lee, AB Whinston (2020) Matching Mobile Applications for Cross-Promotion. *Information Systems Research*, forthcoming.

#N Archak, A Ghose, PG Ipeirotis (2011) Deriving the Pricing Power of Product Features by Mining Consumer Reviews. *Management Science* 57(8): 1485-1509.

#A Ghose, B Li, S Liu (2019) Mobile Targeting Using Customer Trajectory Patterns. *Management Science* 65(11): 5027-5049.

#X Liu, PV Singh, K Srinivasan (2016) A Structured Analysis of Unstructured Big Data by Leveraging Cloud Computing. *Marketing Science* 35(3): 363-388.

#Q Wang, B Li, PV Singh (2018) Copycats vs. Original Mobile Apps: A Machine Learning Copycat-Detection Method and Empirical Analysis. *Information Systems Research* 29(2): 273-291.

#D Puranam, V Narayan, V Kadiyali (2017) The Effect of Calorie Posting Regulation on Consumer Opinion: A Flexible Latent Dirichlet Allocation Model with Informative Priors. *Marketing Science* 36(5): 726-746.

#S Lu, L Xiao, M Ding (2016) A Video-Based Automated Recommender (VAR) System for Garments. *Marketing Science* 35(3): 484-510.

List 11. Economics of AI

Jia, N., Luo, X., Fang, Z., & Liao, C. (2023). When and how artificial intelligence augments employee creativity. *Academy of Management Journal*, (ja).

Lou, B., & Wu, L. (2021). AI ON DRUGS: CAN ARTIFICIAL INTELLIGENCE ACCELERATE DRUG DEVELOPMENT? EVIDENCE FROM A LARGE-SCALE EXAMINATION OF BIO-PHARMA FIRMS. *MIS Quarterly*, *45*(3).

Acemoglu, D., Autor, D., Hazell, J., & Restrepo, P. (2022). Artificial intelligence and jobs: evidence from online vacancies. *Journal of Labor Economics*, *40*(S1), S293-S340.

Felten, E., Raj, M., & Seamans, R. (2021). Occupational, industry, and geographic exposure to artificial intelligence: A novel dataset and its potential uses. *Strategic Management Journal*, *42*(12), 2195-2217.

Brynjolfsson, E., Li, D., & Raymond, L. R. (2023). *Generative AI at work* (No. w31161). National Bureau of Economic Research.

Eisfeldt, A. L., Schubert, G., & Zhang, M. B. (2023). *Generative ai and firm values* (No. w31222). National Bureau of Economic Research.

Agrawal, A., Gans, J., & Goldfarb, A. (2018). *Prediction machines: the simple economics of artificial intelligence*. Harvard Business Press.

List 12. (Static) Discrete Choice Models

A Ghose, PG Ipeirotis, B Li (2012) Designing Ranking Systems for Hotels on Travel Search Engines by Mining User-Generated and Crowdsourced Content. *Marketing Science* 31(3): 493-520.

#A Ghose, SP Han (2014) Estimating Demand for Mobile Applications in the New Economy. *Management Science* 60(6): 1470-1488.

#Y Dong, S Song, S Venkataraman, Y Yao (2020) Mobile Money and Mobile Technologies: A Structural Estimation. *Information Systems Research*, forthcoming.

#B Li, X Li, H Liu (2018) Consumer Preference, Cannibalization, and Competition: Evidence from the Personal Computer Industry. *MIS Quarterly* 42(2): 661-678.

#Z Li, A Agarwal (2017) Platform Integration and Demand Spillovers in Complementary Markets: Evidence

from Facebook's Integration of Instagram. *Management Science* 63(10): 3438-3458.
A Nevo (2000) A Practitioner's Guide to Estimation of Random-Coefficients Logit Models of Demand. *Journal of Economics and Management Strategy* 9(4): 513-548.
S Berry, A Pakes (2007) The Pure Characteristics Demand Model. *International Economic Review* 48(4): 1193-1225.
DW Vincent (2015) The Berry-Levinsohn-Pakes Estimator of the Random-Coefficients Logit Demand Model. *The Stata Journal* 15(3): 854-880.