ECSR, Collegio Carlo Alberto and NASP

Joint Spring School on

New and Old Data:

Cutting-Edge Approaches to Quantitative Social Research

March the 25th - the 29th, 2019

Collegio Carlo Alberto

Piazza Vincenzo Arbarello, 8 – Turin (Italy)

The topic for the seventh edition of the ECSR Spring School is "New and Old Data. Cutting-Edge Approaches to Quantitative Social Research". The School is promoted by the European Consortium of Sociological Research (ECSR), Collegio Carlo Alberto and by the universities of Milan and Turin in the frame of the NASP, Network for the Advancement in Social and Political Studies. It provides high-quality training on the new frontiers of quantitative social research, including computational sociology, network analysis, register data, historical archives data and field experiments. The focus of the school is on the challenges coming from the new data sources and from the related techniques to the standard approach of empirical sociological research, still predominantly based on the estimation of regression-type models on survey data.

The School is organized by Gabriele Ballarino (University of Milan), Fabrizio Bernardi (European University Institute) and Filippo Barbera (Collegio Carlo Alberto and University of Turin).

Lectures will be given and presentations discussed by Guido Alfani (Bocconi University), Gabriele Ballarino, Fabrizio Bernardi, Federico Bianchi (University of Brescia), Gunn Birkelund (University of Oslo), Stefano Cantalini (University of Milan), Diego Gambetta (Collegio Carlo Alberto), Martin Hällsten (University of Stockholm), Krzysztof Krakowski (European University Institute), Nazareno Panichella (University of Milan), Wojtek Przepiorka (University of Utrecht), Flaminio Squazzoni (University of Milan), Arnout van de Rijt (Utrecht University) and Sonja Vogt (University of Bern).

Overview and timetable

	Monday 25 March	Tuesday 26 March	Wednesday 27 March	Thursday 28 March	Friday 29 March
9.30-11.30	*Vogt*	*Birkelund*	*Przepiorka*	*Gambetta*	*Squazzoni*
11.30-11.45	Break	Break	Break	Break	Break
11.45-13.15	*Presentations 1 (Vogt)*	*Presentations 3 (Birkelund)*	*Presentations 5 (Przepiorka)*	*Presentations 7 (Gambetta, Ballarino)*	*Presentations 9 (Squazzoni, Bianchi)*
13.15-14.15	Lunch	Lunch	Lunch	Lunch	Lunch
14.15-16.15	*Hällsten*	*Van de Rijt*	*Krakowski* (14.15-17.15)	*Alfani*	*Bianchi*
16.15-16.30	Break	Break	Break (17.15-17.30)	Break	END
16.30-18.45	*Presentations 2 (Hällsten, Ballarino)*	*Presentations 4 (Van de Rijt, Cantalini)* (16.30-18.00)	*Presentation 6 (Krakowski, Panichella)* (17.30-19.00)	*Presentations 8 (Alfani)* (16.30-18.00)	

1. Keynote lectures

Sonja Vogt: Empirical Methods for Research on Social Development

The course will give an overview of how to design and implement small- and large-scale experiments on sustainability and social development in the field. Specifically, the course will cover the differences and similarities between lab-in-the-field experiments, natural experiments, and randomized controlled trial studies. We will discuss different ways to randomize in the field, ethics, and how to critically link evidence-based field research to policy recommendations on social development. Moreover, the course will cover novel methods on avoiding social desirability biases.

Key references:

Glennerster, R., Takavarasha, K. (2013). *Running randomized evaluations: A practical guide*. Princeton University Press. Chapters 1 & 2.

Martin Hällsten: Everything You Always Wanted to Know About Register Data But Were Afraid to Ask

Register data is increasingly being used for social science research, primarily within Scandinavian countries but also elsewhere. These data has revolutionized the social sciences, but also come with certain limitations. I will discuss the origins of register data, their strengths and limitation in relation to topics such as inference, measurement errors, and types of data and information. I will close with some examples of studies that use register data to their full potential.

Key references:

Almond, D., Edlund, L., Palme, M. (2009). Chernobyl's Subclinical Legacy: Prenatal Exposure to Radioactive Fallout and School Outcomes in Sweden. *The Quarterly Journal of Economics* 124:1729-1772.

Isacsson, G. (2004). Estimating the economic return to educational levels using data on twins. *Journal of Applied Econometrics* 19:99-119.

Korn, E. L., Graubard, B. I. (1998). Variance estimation for superpopulation parameters. *Statistica Sinica* 8:1131-1152.

Wallace, B., Cesarini, D., Lindqvist, E., Östling, R. (2016). Wealth, Health, and Child Development: Evidence from Administrative Data on Swedish Lottery Players. *The Quarterly Journal of Economics* 131:687-738.

Gunn Elisabeth Birkelund: Measuring social mechanisms – by means of register data and field experiments

For decades, various theoretical versions of the concept 'social mechanisms' have been discussed in sociology (e.g., the work of Jon Elster, Peter Hedström and Richard Swedberg), yet empirical studies have shown that it is not an easy task to attempt to measure this concept. In this talk I will discuss two approaches that might be useful illustrations of how one might try to come closer to an empirical understanding of social mechanisms. The first topic is related to ethnic school segregation, and the expected negative impact of minority peers on – in particular – native students' educational outcomes. Theoretically, the literature assumes peer effects to be operating, that is, peers interacting and influencing each other at school. I illustrate this topic by discussing two empirical papers (Hermansen & Birkelund 2015; Branden et al. 2018), which attempt to measure the impact of peers at school on educational outcomes. Second, I will discuss an example of how one might understand a decision making process by developing concepts that could later be tested empirically. The paper used for illustration is Birkelund, Heggebø and Rogstad (2017).

Key references:

Hermansen, A. S., Birkelund, G. E. (2015). The impact of immigrant classmates on educational outcomes. *Social Forces* 94(2), 615- 646.

Brandén, M., Birkelund, G. E., Szulkin, R. (2018). Ethnic Composition of Schools and Students' Educational Outcomes: Evidence from Sweden. *International Migration Review* 1- 32. doi:10.1177/0197918318769314.

Birkelund, G. E., Heggebø, K., Rogstad, J. (2017). Additive or Multiplicative Disadvantage? The Scarring Effects of Unemployment for Ethnic Minorities. *European Sociological Review* 33(1), 17- 29. doi:10.1093/esr/jcw030.

On social mechanisms:

Hedström, P., Swedberg, R. (1998). *Social mechanisms*. Cambridge Univ press.

Elster, J. (2007). *Explaining Social Behavior*. Cambridge Univ press.

Elster, J. (1989). *Nuts and Bolts for the Social Sciences*. Cambridge Univ press.

Arnout van de Rijt: Studying cumulative advantage, Matthew effects, and runaway inequality with new kinds of data and methods

Much work in stratification research focuses on *exogenous* determinants of inequality: Categories such as class, gender, and race differentiate those with more and less socio-economic success. Over the past 7 or so years I have been studying inequalities that *endogenously* form, as the result of

positive feedback operating on random initial differences in success. Such emergent inequality is often referred to as cumulative advantage, Matthew effect, success-breeds-success or rich-get-richer processes. It turns out to be quite tricky to study this phenomenon empirically. I discuss a number of research methods and analysis techniques I and others have used in this endeavor.

Key references

Bol, T., de Vaan, M., van de Rijt, A. (2018). The Matthew effect in science funding. *Proceedings of the National Academy of Sciences*, 115(19), 4887-4890.

Frey, V., Van De Rijt, A. (2016). Arbitrary inequality in reputation systems. *Scientific reports*, 6, 38304.

Restivo, M., Van De Rijt, A. (2012). Experimental study of informal rewards in peer production. *PloS one*, 7(3), e34358.

Salganik, M. J., Dodds, P. S., Watts, D. J. (2006). Experimental study of inequality and unpredictability in an artificial cultural market. *Science*, 311(5762), 854-856.

Van de Rijt, A., Kang, S. M., Restivo, M., & Patil, A. (2014). Field experiments of success-breeds-success dynamics. *Proceedings of the National Academy of Sciences*, 111(19), 6934-6939.

Wojtek Przepiorka: Conducting field experiments online and offline: Design, implementation and analysis

Behavioral laboratory experiments facilitate the identification of causal mechanisms and cumulative research because of the highly controllable context in which they take place. However, lab experiments have been criticized for the lack of realism, especially when claims about the validity of their results are made beyond the lab context. Field experiments therefore seem suitable to extend the validity of lab findings into the "real" world. However, field experiments are more difficult to conduct just because they take place in contexts that are not highly controllable. In this lecture, I will talk about the practical issues that can arise with regard to design, implementation and analysis of field experiments. I will thereby draw on my experience with conducting a field experiments on ethnic discrimination in online markets and the enforcement of social norms in public transport.

Key references:

Preparatory readings

Baldassarri, Delia, and Maria Abascal. 2017. "Field Experiments Across the Social Sciences." *Annual Review of Sociology* 43:41-73.

Parigi, Paolo, Jessica J. Santana, and Karen S. Cook. 2017. "Online Field Experiments: Studying Social Interactions in Context." *Social Psychology Quarterly* 80(1):1-19.

Further readings

Balafoutas, Loukas, and Nikos Nikiforakis. 2012. "Norm enforcement in the city: A natural field experiment." *European Economic Review* 56(8):1773-85.

Berger, Joël, and Debra Hevenstone. 2014. "Norm enforcement in the city revisited: An international field experiment of altruistic punishment, norm maintenance, and broken windows." *Rationality and Society* 28(3):299-319.

Doleac, Jennifer L., and Luke C.D. Stein. 2013. "The Visible Hand: Race and Online Market Outcomes." *Economic Journal* 123(572):F469-F92.

Gerber, Alan S., and Donald P. Green. 2012. *Field Experiments: Design, Analysis, and Interpretation*. New York: W. W. Norton & Company.

Harrison, Glenn W., and John A. List. 2004. "Field Experiments." *Journal of Economic Literature* 42(4):1009–55.

Keizer, Kees, Siegwart Lindenberg, and Linda Steg. 2008. "The Spreading of Disorder." *Science* 322(5908):1681-85.

Pager, Devah, Bruce Western, and Bart Bonikowski. 2009. "Discrimination in a Low-Wage Labor Market: A Field Experiment." *American Sociological Review* 74(5):777-99.

Przepiorka, Wojtek. 2011. "Ethnic Discrimination and Signals of Trustworthiness in Anonymous Online Markets: Evidence from Two Field Experiments." *Zeitschrift für Soziologie* 40(2):132-41.

Przepiorka, Wojtek, and Joël Berger. 2016. "The sanctioning dilemma: A quasi-experiment on social norm enforcement in the train." *European Sociological Review* 32(3):439-51.

van de Rijt, Arnout, Soong Moon Kang, Michael Restivo, and Akshay Patil. 2014. "Field experiments of success-breeds-success dynamics." *Proceedings of the National Academy of Sciences of the U.S.A.* 111(19):6934-39.

Krzysztof Krakowski: Using non-conventional data for quantitative analysis

The session discusses the use of non-conventional data for quantitative analysis, focusing on applications across many fields in social sciences (sociology, political science, economics). The session is divided in two parts. In the first part, we discuss main advantages of using non-conventional quantitative data vis-à-vis using survey data. We then review the potential sources of nonconventional data, such as historical archives, social media content, or satellite imagery. Illustrations of successful applications from the literature are provided throughout the session. In the second part, we focus on a potential use of nonconventional quantitative data in the participants' own research projects. The

participants are therefore encouraged to share abstracts of their projects with the instructor before the session (email: krzysztof.krakowski@carloalberto.org).

Key references:

Fisman, R., Miguel, E. (2010). *Economic gangsters: corruption, violence, and the poverty of nations*. Princeton University Press.

Lee, M. M., Zhang, N. (2017). Legibility and the informational foundations of state capacity. *The Journal of Politics*, 79(1), 118-132.

Krakowski, K., A. Rink, M. Schaub. (2018). *Ethnic Riot and Prosocial Behavior: Evidence from Kyrgyzstan*. Working Paper.

Barceló, J., Labzina, E. (forthcoming), Do Islamic State's Deadly Attacks Demotivate, Deter, or Mobilize Supporters? *British Journal of Political Science*.

Diego Gambetta: What can we do if "causal identification" is unattainable?

Many important research questions in the social sciences cannot be answered experimentally or with some other causally robust identification method. In this class I consider some alternatives ways of approaching research questions of this kind, which we can generically dub "implication analyses". I consider different routes to drawing implications: starting from a theory or from a proven mechanism, or from some combination thereof. The generic form of reasoning is simple: if the theory or mechanism were the cause of X, then we should expect to find that Y1 obtains and Y2 does not. Implication analyses is often good to rule causes out, which while not perfect is a step forward. I also discuss how puzzles, anomalous or unexpected correlations, can help up restrict the set of possible theories and sharpen the implications we can draw. As examples I will use suicide, extremism, crime decline and others. Implications that could be useful to your research can be discussed in class.

Key references:

Durkheim E. (1897). *Suicide.* [The book is a great read. It relies on implications which Durkheim derives from various conjectures on what might be the causes of suicide; read especially chapter 4, on *Imitation*] [In the original French the book can be found here] http://www.gutenberg.org/ebooks/40489; in English here http://archive.org/details/suicidestudyinso00durk]

Gambetta, D., Hertog, S. (2016). *Engineers of Jihad. The curious connection between violent extremism and education*. Princeton University Press, chapters 3 and 4 especially. [Here the range

of theories is narrowed through a puzzle, and different implications are drawn from the remaining theories]

Batuman, E. (2013). "Poisoned Land. On the trail of a mystery disease in the Balkans". *New Yorker*, August 12, 2013 [the search for the causes of a potentially fatal kidney disease, Balkan endemic nephropathy] https://www.newyorker.com/magazine/2013/08/12/poisoned-land

Wolpaw Reyes, J. (2007). *Environmental Policy as Social Policy? The Impact of Childhood Lead Exposure on Crime*. NBER WP No. 13097 www.nber.org/papers/w13097.pdfwww.nber.org/papers/w13097.pdf

Those with abstract methodological tastes can find some useful arguments in the next two references (which are short of examples):

Lieberson, S., Lynn, F. B. (2002). Barking up the wrong tree. Scientific Alternatives to the Current Model of Sociological Research. *Annual Review of Sociology* 28:1–19.

Lieberson, S., Horwich, J. (2008). Implications analysis: a pragmatic proposal for linking theory and data in the social sciences. *Sociological Methodology* 38(1): 1-50. [discussed by several authors here: http://onlinelibrary.wiley.com/doi/10.1111/some.2008.38.issue-1/issuetoc]

Guido Alfani: Economic inequality and social mobility in preindustrial Europe

Recent research in economic history has unearthed previously unknown facts about the long-term trends in inequality. We now have, for at least some areas of Europe, continuous time series of key inequality indicators from ca. 1300. Most of these series have resulted from the research conducted by the ERC-funded project *EINITE – Economic Inequality across Italy and Europe 1300-1800*. These new data are changing the way in which we perceive economic inequality not only in the past, but even today – as a key lesson from history, is that economic inequality (especially, but not only, of wealth) has a marked tendency for increasing over time, and only catastrophes on the scale of the Black Death or the World Wars managed to bring it down, albeit temporarily. Additionally, the new historical evidence is also relevant to the debate about the long-term determinants of inequality growth. This seems to be independent, to a large degree at least, from economic growth. Other factors seem to have played a crucial role, including institutional factors and in particular (in the early modern period) the rise of the fiscal-military state. These recent acquisitions, however, raise many questions about the actual impact on society of distributive dynamics. A more recent ERC project, *SMITE – Social Mobility and Inequality across Italy and Europe 1300-1800*, is currently exploring at least some key aspects of the social impact and significance of inequality change. The seminar, based on case studies from different European areas and especially from Italy and the Low Countries, will provide an overview of the research done or underway in the context of the aforementioned projects.

Key references:

Alfani, G. (2017), "The rich in historical perspective. Evidence for preindustrial Europe (ca. 1300-1800)", *Cliometrica*, 11(3): 321-348.

Alfani, G. and W. Ryckbosch (2016), "Growing apart in early modern Europe? A comparison of inequality trends in Italy and the Low Countries, 1500–1800", *Explorations in Economic History*, 62, 143-153.

Alfani, G. and M. Di Tullio (2019), The Lion's Share. Inequality and the Rise of the Fiscal State in Preindustrial Europe, Cambridge University Press (chapter 4).

Flaminio Squazzoni and Federico Bianchi: Social network analysis: from data collection to statistical modelling

The lecture aims to introduce students to empirical research on social network. By allowing social scientists to represent actors and their relationships as nodes and edges of a network, social network analysis is key to explain aggregate patterns and dynamics in a variety of social contexts, from markets to complex organizations. First, some key results of social network research will be covered and discussed. Students will then be introduced to the basic designs for empirical network data collection, focusing on sociocentric (full-network) design. After a brief introduction to basic R commands, students will then familiarize with the statnet-family packages in R, which will be used to manage a network dataset and calculate basic descriptive statistics. Finally, students will test hypotheses about formation mechanisms of an empirical network by estimating basic statistical network models (Exponential Random Graph Models).

Students will practically work in R with empirical data provided by the instructor and based on: Bianchi, F., Casnici, N., & Squazzoni, F. 2018. "Solidarity as a byproduct of professional collaboration. Social support and trust in a cowering space." *Social Networks* 54: 61—72. doi: 10.1016/j.socnet.2017.12.002 (paywalled article, available upon request to the instructors).

Students will be required to have R and RStudio already installed in their laptops (free softwares):
https://cran.r-project.org/
https://www.rstudio.com/products/rstudio/download/#download

Key references:

Robbins, G. (2015). *Doing Social Network Research. Network-Based Research Designs for Social Scientists*. London: Sage.

Borgatti, S.P, Everett, M.G., & Johnson, J.C. (2013). *Analyzing Social Networks*. London: Sage.

Free introductory course to R: https://www.datacamp.com/courses/free-introduction-to-r

Students' presentations

Day	Presentation session	Authors and discussants		
Monday, 25/3	1	C. J. Gil Hernández, *Birth Weight and Early School Readiness: Biological Destiny or Parental Response?* Discussant: S. Vogt	J. J. Kuhnle, *Does Deregulation Lead to Better Economic Performance Among Immigrants? The 2004 Reform of the German Trade and Crafts Code as a Natural Experiment* Discussant: S. Vogt	
	2	L. Azzollini, *The Lost and Scarred Generation: Income Inequality, Social Stratification, and Turnout Inequality in Europe* Discussant: G. Ballarino	M. Dardoumpa, *The differences in perinatal health between migrant and native children* Discussant: M. Hällsten	G. Scarano, *Do payment incentives work to avoid opportunistic behavior for quasi-markets in employment services? Evidence from Italy at the dawn of minimum income reform* Discussant: M. Hällsten
Tuesday, 26/3	3	G. Piccitto, *Children occupational status & career and the menthal health of their parents* Discussant: G. E. Birkelund	S. Skovgaard Jensen, *School Wide Positive Behaviour Support for problems at school* Discussant: G. E. Birkelund	
	4	I. Plavgo, *Education for all, completion for some? Trends and determinants of intergenerational educational inequality in sub-Saharan Africa* Discussant: A. van de Rijt	V. J. Deimantas, *Positive Relationship Between Migration and Fertility Intentions: the Case of Postsocialist EU Member States* Discussant: S. Cantalini	

Wednesday, 27/3	5	S. Ramkumar, *Importance of network consequences in actors' decisions to engage in circular, closed loop innovation – a case study of REALCAR* Discussant: W. Przepiorka	E. Shliakhovchuk, *"Empathy Video Games": Video Games as Awareness Raisers, Attitude Changers and Agents of Social Change* Discussant: W. Przepiorka	
	6	M. Chen, *Dating webs and marriage in China* Discussant: K. Krakowski	M. Cozzani, *Prenatal Pollution, Children Development, and Parental Response* Discussant: N. Panichella	
Thursday, 28/3	7	J. Beckmann, *Does psychological counselling reduce the gender gap in field of study aspirations? Evidence from randomized controlled trial* Discussant: D. Gambetta	J. I. Castillo Jaramillo, *The Impact of Interviewer Effects on Skin Color Assessment in a Cross-National Context* Discussant: G. Ballarino	
	8	G. Corti, *Marriage choices and Social Homogamy through 19th and 20th centuries: the case of Milan* Discussant: G. Alfani	R. Molinari, *The institutionalisation of death in hospital settings in Italy: an historical perspective (1883-2013)* Discussant: G. Alfani	
Friday, 29/3	9	A. Taalaibekova, *Who matters in coordination problems on networks: myopic or farsighted agents?* Discussant: F. Squazzoni	R. Topinková, *It Takes Two to Tango: Popularity on a mobile dating app* Discussant: F. Bianchi	