



ABSTRACT BOOKLET

SESSION 1: Micro-level Data I

Leigh Shaw-Taylor, *University of Cambridge*

Title: The Occupational Structure of England and Wales 1379-1911 and some reflections on: Data (not necessarily 'big'), the 'hypothesis first' dogma, questions and the origins of hypotheses

Abstract:

Ewen Gallic, *Aix-Marseille School of Economics*

Title: Modeling Joint Lives within Families

Co-authors: Olivier Cabrignac (SCOR) and Arthur Charpentier (UQAM)

Abstract: Family history is usually seen as a significant factor insurance companies look at when applying for a life insurance policy. Where it is used, family history of cardiovascular diseases, death by cancer, or family history of high blood pressure and diabetes could result in higher premiums or no coverage at all. In this article, we use massive (historical) data to study dependencies between life length within families. If joint life contracts (between a husband and a wife) have been long studied in actuarial literature, little is known about child and parents dependencies. We illustrate those dependencies using 19th century family trees in France, and quantify implications in annuities computations. For parents and children, we observe a modest but significant positive association between life lengths. It yields different estimates for remaining life expectancy, present values of annuities, or whole life insurance guarantee, given information about the parents (such as the number of parents alive). A similar but weaker pattern is observed when using information on grandparents. Thus, it seems that most of the information insurers expect to learn from family history could probably be captured by other variables that could capture these common environmental factors that individuals share with their ancestors.

SESSION 2: Micro-level Data II

Steven Ruggles, *University of Minnesota*

Title: Big Data: Data Infrastructure and the Paranoid Style in the United States, 1965-2020

Abstract:

Katherine A Eriksson, *University of California, Davis*

Title: Immigration and Infant Mortality in Massachusetts: Evidence from the Age of Mass Migration

Co-authors: Peter Z. Lin (University of California, Davis) and Greg Niemesh (Miami University)

Abstract: This paper estimates the effect of immigration on infant mortality during the Age of Mass Migration. Specifically, we use a shift-share instrument and town-level panel data from Massachusetts between 1860 and 1915 to estimate the impact over a long period. We find a significant positive effect of immigrant inflows on native infant mortality throughout the whole period. We also find suggestive evidence that this effect is due to communicable diseases and over-crowding. Public health investments do not appear to have mitigated the effect.

Shari Eli, *University of Toronto*

Title: Do Youth Employment Programs Work? Evidence from the New Deal

Co-authors: Anna Aizer (Brown University), Adriana Lleras-Muney (UCLA) and Keyoung Lee (Yale University)

Abstract: We study the Civilian Conservation Corps (CCC) – the first and largest youth training program in the U.S. in operation between 1933 and 1942 – to provide the first comprehensive assessment of the short- and long-term effects of means-tested youth employment programs. We use digitized enrollee records from the CCC program in Colorado and New Mexico and matched these records to the 1940 Census, WWII enlistment records, Social Security Administration records, and death certificates. We find that enrollees who spent more time in CCC training grew taller, lived longer lives and had higher lifetime earnings as a result of their participation in the program. We also find modest increases in the educational attainment of the participants and increases in short term geographic mobility. In contrast, we find no evidence that their labor force participation or wages increased in the short run. To assess the internal and external validity of the results, we compare our estimates to those derived from a randomized evaluation of Job Corps, the modern version of the CCC, conducted in the 1990s. The RCT's results show that our empirical strategy delivers estimates that are in line with the experimental estimates. Overall, we find significant long-term benefits in both longevity and earnings, suggesting short and medium-term evaluations underestimate the returns of training programs, as do those that fail to consider effects on longevity.

SESSION 3: Textual Data

Michela Giorcelli, *University of California, Los Angeles*

Title: Technology Transfer and Early Industrial Development: Evidence from the Sino-Soviet Alliance

Co-authors: Bo Li (Tsinghua University)

Abstract: This paper studies the causal effect of technology transfer on early industrial development. Between 1950 and 1957, the Soviet Union supported the “156 Projects” in China for the construction of technologically advanced, large-scale, capital-intensive industrial facilities. We rely on idiosyncratic delays in project completion combined with the Sino-Soviet Split, due to which some of them received Soviet technology and know-how, while some others were eventually realized by China alone. We find that the Soviet technology had large and persistent effects on plant performance, and know-how diffusion by Soviet experts further boosted firm outcomes. The intervention generated horizontal and vertical spillovers and production reallocation from state-owned to privately-owned companies since the late 1990s.

Elliott Ash, *University of Zurich*

Title: Emotion and Reason in Political Language

Co-authors: Gloria Gennaro (University of Zurich)

Abstract: We use computational linguistics techniques to study the use of emotion and reason in political discourse. Our new measure of emotionality in language combines lexicons for affective and cognitive processes, as well as word embeddings, to construct a dimension in language space between emotion and reason. After validating the method against human annotations, we apply it to scale 6 million speeches in the U.S. Congressional Record for the years 1858 through 2014. Intuitively, emotionality spikes during time of war and is highest for patriotism-related topics. In the time series, emotionality was relatively low and stable in the earlier years but increased significantly starting in the late 1970s. Comparing Members of Congress to their colleagues, we find that emotionality is higher for Democrats, for women, for ethnic/religious minorities, and for those with relatively extreme policy preferences (either left-wing or right-wing) as measured by roll call votes.

SESSION 4: Micro-level Data III

Lionel Kesztenbaum, *National Institute for Demographic Studies, INED*

Title: Building a Historical Matrix. Lessons from the TRA Database

Abstract:

Guillaume Blanc, *Brown University*

Title: Demographic Change and Development from Crowdsourced Genealogies in Early Modern Europe

Abstract: This paper draws on a novel historical dataset crowdsourced from publicly available genealogies to study demographic change and development at the individual level in the distant past. I reconstruct fertility series, identify migration in and out of urban centers, and provide novel measures and stylized facts in a period without census and with millions of ordinary individuals observed in thirty countries. For each country, I carefully show that selection is limited in the data. Then, I document patterns of human mobility, fertility, and adult mortality in Early Modern Europe, through the Industrial Revolution and demographic transition. Finally, I present several findings at a disaggregated level suggesting that substantial and rapid changes in preferences took hold with the Age of Enlightenment and played an important role in the transition from stagnation to growth. In particular, I estimate the onset of the decline in fertility in France in the 1760s, a hundred years before the rest of Europe and earlier than previously thought, and I find a weaker intergenerational persistence of fertility behavior in Europe as early as in the late eighteenth century.

SESSION 5: Pre-Modern Data

Mattia Fochesato, *Bocconi University*

Title: “Pandemics and Socio-economic Status. Evidence from the Plague of 1630 in Northern Italy”

Co-authors: Guidi Alfani and Marco Bonetti (Bocconi University)

Abstract: This article provides a unique analysis of the biological, socio-economic and institutional factors shaping the individual risk of death during a major preindustrial plague. We use a deep microdemographic database for the city of Carmagnola in North Italy during the 1630 plague. An exceptional feature of this database is that it allows us to explore in detail the survival dynamics of the population hospitalized at the plague ward (the *lazzaretto*). We develop a theoretical model of the admissions to the *lazzaretto* to better understand the observational data, which we then analyse by means of econometric techniques. We explore the impact of biological factors (age and sex) in shaping the individual risk of death but more importantly, we provide a one-of-a-kind study of the impact of social-economic status, a variable notoriously difficult to measure with precision for pretransitional societies. We report an inversion of the normal mortality gradient by status for those interned at the *lazzaretto*, and interpret this as the consequence of a greater ability of the rich to make decisions about their care, an ability which ultimately backfired. At the other extreme of the social ladder, the poor admitted to the *lazzaretto* suffered a relatively low risk of death because they enjoyed much better conditions that they would have had outside the hospital and because they had probably been the object of some preventive internment, allowing them to avoid infection with relative ease.

Yossef Rapoport, *Queen Mary University London*

Title: “Quantitative and GIS Analysis of a Rural Tax Register from Medieval Egypt (1245 CE)”

Abstract: *The Villages of the Fayyum* is a record of an audit of over one hundred villages in the Egyptian province of the Fayyum, composed in 1245 CE by the Cairene bureaucrat Abū ‘Uthmān al-Nābulusī’s. Preserved in manuscript as part of a historical and geographical account of the province, it is the most detailed surviving tax register of any region in the medieval Islamic world.

The records for each village include information on 1) size and location, tribal affiliation, religious composition, water sources and the state-appointed recipient of the village’s fiscal revenue 2) amounts of agricultural and commercial taxes and fees in kind and in cash, by crop 3) Credits and seed advances 4) Poll-tax on non-Muslims 5) State-owned sugar-cane plantations and presses 6) Horse-riders for military campaigns.

The location of most villages can be reconstructed through their modern locations. Further historical context is provided by sets of documents unearthed in deserted villages of the Fayyum, which go back to Roman, Byzantine and earlier Islamic periods.

Quantitative and GIS analysis of the fiscal and demographic data shows clear correlation between grain production on the one hand, and tribal and Muslim identity on the other. This finding has significant implications for understanding the process of the conversion of rural Egypt from Christianity to Islam. The data also enables a refined reconstruction of local economy, such as the relationship between grains and small cattle in a mixed economy, and the sugar-cane network.

SESSION 6: Georeferenced Data

Eric Melander, *University of Namur*

Title: Transportation Technology, Individual Mobility and Social Mobilisation

Abstract: Between 1881 and 1910, Swedish society underwent two transformative developments: the large-scale roll-out of a national railway network and the nascence of grassroots social movements which came to dominate economic, social and political spheres well into the twentieth century. Using exogenous variation in railway access arising from initial plans for the network, I show that as localities became better-connected, they were more likely to host a local movement and saw more rapid membership growth and a greater number of distinct organisations. The mobility of individuals is a key mechanism: results are driven by passenger arrivals into connected localities, not freight arrivals. I implement a market access framework to show that, by reducing least-cost distances between localities, railways intensified the influence exerted by neighbouring concentrations of membership, thereby enabling social movement spread. Subsequently – in Sweden’s first election with universal male suffrage in 1911 – localities with greater social movement mobilisation exhibited higher turnout and Social Democrat vote shares.

Thomas Thévenin, *Université de Bourgogne* and **Alexis Litvine**, *University of Cambridge/CAMPOP*

Title: ANR-Communes: a Big Geospatial Data Infrastructure as a Gateway to Even Bigger Data

Abstract: We will present the data infrastructure we are developing for our project ANR-COMMUNES, which is both the result of a very ambitious campaign of data collection, digitalization and structuration and, critically, the gateway to an integration of even bigger historical and geospatial datasets through the new H-GIS platform we are creating. In this presentation, we will first explain our methodology to acquire and manage the very large data load required to build this infrastructure. By creating annual boundary data and transport links for over 40,000 territorial units over two centuries we are operating in a multidimensional environment, associating for each point in space and for each year multiple Origin-Destination matrices (one per transport mode, and several multi-modal options depending on commodities transported and parameters adopted), each one containing c.1.6 billion pairs. We will then explain our vision and key methodological principle for our project, which is that all data (historical and modern, both existing and yet to be created) should be linked to this common H-GIS infrastructure at the highest level of spatial disaggregation possible, using an open and interoperable architecture (linked data/linked places). Using data extracted from conscription listings, we will conclude this presentation by briefly demonstrating how social and economic datasets can be integrated and enhanced by our data infrastructure.