New Administrative Data and New Research Designs in Economics

Miklós Koren (CEU) “Challenges of Multidimensional Transactional Data”

Caterina Calsamiglia (CEMFI), “Administrative Records and the Metamorphosis of Economics of Education”


Andrea Weber (Vienna) "Matched Individual Careers Across Administrative Registers of Two Countries"

Chair Richard Blundell.

Summary of Presentations

Miklós Koren “Challenges of Multidimensional Transactional Data”

Recent empirical work has used customs transactions data to analyze the patterns of international trade. The availability of such data has opened up the possibility to ask questions beyond the volume of trade and its broad sectoral composition. A typical customs declaration (which serves as the primary unit of observation for most trade statistics) records the exporting and the importing firm, the precise classification of the product being shipped, the precise date of shipment, the mode of transport, and many other logistical details about shipment. This has made it possible, for example, to study the distribution of trade across products, destination markets and firms.

We note that customs data is very similar to transactional data collected by many businesses. Each transaction is characterized by many categorical variables (“dimensions”) and few quantitative variables (“measures” or “facts”). The data can then be effectively represented by a star schema, where each fact is linked to its many dimensions. Such data can be used to study retail purchases, business supply transactions, or international and domestic shipments, for example.

The proliferation of the number of dimensions poses a conceptual as well as a computational challenge to empirical research. The conceptual challenge relates to the sparsity of such data, that is, the high number of potential categories relative to the number of observations. For example, consider a customs dataset recording the exact date of exports, one of 365 days of the year. The potential number of different classifications (firms, products, destination countries and days of the year) can easily mount to 100 trillion categories. It is not only computationally, but also conceptually infeasible to treat these categories as independent observations representing an underlying model of international trade. The computational challenge relates to the need of controlling for unobserved heterogeneity via fixed or random effects. Using standard within
transformation is simply not feasible with so many dimension, and neither is the practice of estimating the fixed effects directly.

The talk highlights several new empirical approaches to working with multidimensional data with a large number of dimensions.

Caterina Calsamiglia “Administrative Records and the Metamorphosis of Economics of Education”

A summary of how the landscape for Economics of Education has changed thanks to the explosion in the access of administrative records from a rich set of countries. The presentation will briefly discuss how we have gone from measuring returns to education from cross section data that looked at years of education and earnings to being able to exploit a reform in the expansion of compulsory education to get at better identified returns, to estimates using a large panels from Norway to compare the estimates from panel analysis, the expansion of compulsory education and differences in month of birth, all at once. The question is very old, but only in the recent years have we been able to provide some reliable answers.

More data and simple but important questions being addressed has only led to an explosion in the number of studies, techniques to be used and data to be exploited as we have also become more aware of the complexity or richness of the answers to those questions: life cycle components of these answers, heterogeneous effects, poor measurements of attainment (more so the more exams are being used for policy ...more teaching to the test) of student and teachers, etc... An explosion of questions being addressed but also of unanswered questions arising!

Finally the talk will discuss school choice more specifically, to line out the developments: from Hoxby exploiting Tiebout choice a decade ago to the big emphasis on the details of the mechanisms implemented, to slowly the merger of the two approaches to offer a more complete and complex story of what there is to understand when drawing conclusions on school choice. All this with the help of both theory, rich data, and novel econometric approaches to identify parents’ and governments’ preferences.

Magne Mogstad “Opportunities and Challenges of Using Administrative Data in Empirical Research”

The term big data is currently a buzzword in social science. Some refer to administrative data as big data. Administrative data are defined as data which derive from the operation of administrative systems, typically by public sector agencies. While not collected for research purposes, administrative data offers new opportunities and challenges for social science. In this talk, I give a few examples of recent empirical research exploiting two important opportunities offered by administrative data. First, administrative data typically contain information about the entire population. This allows researchers to study particular subgroups without having to worry as much about small sample sizes
or non-sampling errors; sometimes these groups are of intrinsic interest; other times they are key to understand market level outcomes. Second, administrative registers typically have common unit identifiers, making it possible to link various data sets which is necessary for certain analyses."

Andrea Weber "Matched Individual Careers Across Administrative Registers of Two Countries"

In many countries administrative data have become available for search, allowing to follow detailed careers of individuals over their lifetimes. However, little is known about the individuals who they leave the registers of a specific country. Free movement of workers is one of the EU's key principles and it is also at the forefront of the public debate. To provide novel empirical evidence on the movement of workers, we have started a project with the aim of linking Austrian and German social security registers. The matched register data capture individual movements across the two countries and allow tracing their careers on both sides of the border. In ongoing research we shed light on motives of initial migration, as well as return and circular migration and study the influence of changes in the institutional environment of one country and differences in business cycles on mobility of workers. The long-term goal of the project is to include multiple European countries and form a detailed picture of migration flows across Europe.