

Economists Learn to Grapple with Big Data

BY JESSIE ROMERO AND AARON STEELMAN

Most of us have had the experience of shopping online and receiving a recommendation for an item that caught our eye — and wondered how the suggestion was generated. You have “big data” to thank for it. If you go on to make the additional purchase, so does the retailer. And increasingly, it isn’t just companies that are interested in big data; economists are too.

Our browsing and purchasing experiences, along with those of other consumers with similar tastes and interests, generate an enormous amount of information that can be collected and filtered to provide us with something like a tailor-made shopping environment. As gathering that information has gotten easier and analyzing it has gotten cheaper, businesses are aiming to use it to boost sales. The recommendations we receive may not always be hits, but they don’t have to be. In baseball, a batter who is successful just 30 percent of the time is often an all-star — and most retailers would be happy with an average much lower than that.

What exactly is “big data”? There is no consensus definition, in part because big data is relatively new and in part because people use it for many different purposes. But for economists who work with big data, there is broad agreement on what features make it valuable. Linnet Taylor of the University of Amsterdam and Ralph Schroeder and Eric Meyer of the Oxford Internet Institute have surveyed more than 125 social scientists on various issues related to big data. In a 2014 article published in the journal *Big Data & Society*, they reported that the economists they have talked to are most interested in “granular, population-level data with multiple dimensions that allow researchers to analyse cases along many variables,” which permit them “to test theories of behaviour that were previously untestable, creating a new set of metrics for issues of economic interest which were previously in the realm of theory.”

Economists have worked with big data, both public and private, in almost all areas of their discipline, from historical tax data to look at economic inequality to Medicare rolls to examine the efficiency of the health care industry. But if there is one area where big data may be especially promising, it is labor economics. Indeed, John Horton and Prasanna Tambe of New York University’s Stern School of Business recently noted that we are “clearly entering a golden age for empirical labor market research,” one where there is “a growing opportunity to revisit old questions with new and better data and to answer new questions raised directly by these new contexts.”

Horton and Tambe point to several recent papers that have employed big data to interesting effect. For instance, Iona Marinescu of the University of Chicago and Roland

Rathelot of the University of Warwick used ZIP code level data from 500,000 job seekers who sent out more than 5 million applications through CareerBuilder.com in 2012 to examine the extent to which geographic mismatch is a driver of unemployment. They found that job seekers are 35 percent less likely to apply for a job 10 miles away from their ZIP code of residence — but because there are enough job openings on average, this local preference has been fairly unimportant in the aggregate. In a similar vein, Scott Baker of Northwestern University’s Kellogg School of Management and Andrey Fradkin, a postdoctoral associate at MIT’s Sloan School of Management, used Google search data to look at online job search patterns in Texas and how those patterns change as people come close to exhausting their unemployment insurance benefits.

Of interest to macroeconomists, including monetary policymakers, is the Billion Prices Project at MIT, which collects price data daily from hundreds of online retailers. Those data are used for a variety of research purposes, perhaps most notably to construct a complementary inflation measure to the Consumer Price Index. Such data are collected internationally, too, and could be particularly helpful for countries that do not have as reliable government measures.

Most economists are optimistic about the use of big data in academic research and in the evaluation of public policy. At the same time, most also agree that the size and complexity of some of those data sets will require new statistical techniques to get beyond mere correlation and to the identification of causal relationships that help us test theory. As Jonathan Levin of Stanford University notes, “Everywhere you look you can generate an interesting fact. But figuring out how to turn that into ... a researchable question is really challenging.”

Privacy concerns will loom large, too, as researchers avail themselves of data sets containing sensitive information. “De-identification” methods will need to be robust to ensure appropriate anonymity. Moreover, the proper use of predictive modeling to achieve public policy ends will need to be determined. Big data could be helpful, suggest Levin and his colleague Liran Einav, in helping the government identify people with a high marginal propensity to consume — people who could then be targeted for tax rebates as part of an “economic stimulus” package. Private firms routinely engage in similar activities, of course, but people’s reaction to such measures by the public sector likely would be more circumspect.

What is not in doubt is that big data is going to keep getting bigger. As it does, economists will have to figure out, often in collaboration with colleagues in other disciplines, such as computer science, how to make the best use of it. **EF**