April 28, 2014

New Data Sources: A Conversation with Google's Hal Varian

New Data Sources: A Conversation with Google's Hal Varian

In recent years, there has been an explosion of new data coming from places like Google, Facebook, and Twitter. Economists and central bankers have begun to realize that these data may provide valuable insights into the economy that inform and improve the decisions made by policy makers.

As chief economist at Google and emeritus professor at UC Berkeley, Hal Varian is uniquely qualified to discuss the issues surrounding these new data sources. Last week he was kind enough to take some time out of his schedule to answer a few questions about these data, the benefits of using them, and their limitations.

Mark Curtis: *You've argued that new data sources from Google can improve our ability to "nowcast." Can you describe what this means and how the exorbitant amount of data that Google collects can be used to better understand the present?*

Hal Varian: The simplest definition of "nowcasting" is "contemporaneous forecasting," though I do agree with David Hendry that this definition is probably *too* simple. Over the past decade or so, firms have spent billions of dollars to set up real-time data warehouses that track business metrics on a daily level. These metrics could include retail sales (like Wal-Mart and Target), package delivery (UPS and FedEx), credit card expenditure (MasterCard's SpendingPulse), employment (Intuit's small business employment index), and many other economically relevant measures. We have worked primarily with Google data, because it's what we have available, but there are lots of other sources.

Curtis: *The ability to "nowcast" is also crucially important to the Fed. In his December press conference, former Fed Chairman Ben Bernanke stated that the Fed may have been slow to acknowledge the crisis in part due to deficient real-time information. Do you believe that new data sources such as Google search data might be able to improve the Fed's understanding of where the economy is and where it is going?*

Varian: Yes, I think that this is definitely a possibility. The real-time data sources mentioned above are a good starting point. Google data seems to be helpful in getting real-time estimates of initial claims for unemployment benefits, housing sales, and loan modification, among other things.

Curtis: *Janet Yellen stated in her first press conference as Fed Chair that the Fed should use other labor market indicators beyond the unemployment rate when measuring the health of labor markets. (The Atlanta Fed publishes a labor market spider chart incorporating a variety of indicators.) Are there particular indicators that Google produces that could be useful in this regard?*

Varian: Absolutely. Queries related to job search seem to be indicative of labor market activity. Interestingly, queries having to do with killing time also seem to be correlated with unemployment measures!

Curtis: *What are the downsides or potential pitfalls of using these types of new data sources?*

Varian: First, the real measures—like credit card spending—are probably more indicative of actual outcomes than search data. Search is about intention, and spending is about transactions. Second, there can be feedback from news media and the like that may distort the intention measures. A headline story about a jump in unemployment can stimulate a lot of "unemployment rate" searches, so you have to be careful about how you interpret the data. Third, we've only had one recession since Google has been available, and it was pretty clearly a financially driven recession. But there are other kinds of recessions having to do with supply shocks, like energy prices, or monetary policy, as in the early 1980s. So we need to be careful about generalizing too broadly from this one example.

Curtis: *Given the predominance of new data coming from Google, Twitter, and Facebook, do you think that this will limit, or even make obsolete, the role of traditional government statistical agencies such as Census Bureau and the Bureau of Labor Statistics in the future? If not, do you believe there is the potential for collaboration between these agencies and companies such as Google?*

Varian: The government statistical agencies are the gold standard for data collection. It is likely that real-time data can be helpful in providing leading indicators for the standard metrics, and supplementing them in various ways, but I think it is highly unlikely that they will replace them. I hope that the private and public sector can work together in fruitful ways to exploit new sources of real-time data in ways that are mutually beneficial.

Curtis: *A few years ago, former Fed Chairman Bernanke challenged researchers when he said, "Do we need new measures of expectations or new surveys? Information on the price expectations of businesses—who are, after all, the price setters in the first instance—as well as information on nominal wage expectations is particularly scarce." Do data from Google have the potential to fill this need?*

Varian: We have a new product called Google Consumer Surveys that can be used to survey a broad audience of consumers. We don't have ways to go after specific audiences such as business managers or workers looking for jobs. But I wouldn't rule that out in the future.

Curtis: *MIT recently introduced a big-data measure of inflation called the [Billion Prices Project](). Can you see a big future in big data as a measure of inflation?*

Varian: Yes, I think so. I know there are also projects looking at supermarket scanner data and the like. One difficulty with online data is that it leaves out gasoline, electricity, housing, large consumer durables, and other categories of consumption. On the other hand, it is quite good for discretionary consumer spending. So I think that online price surveys will enable inexpensive ways to gather *certain* sorts of price data, but it certainly won't replace existing methods.

By Mark Curtis, a visiting scholar in the Atlanta Fed's research department

April 28, 2014 in [Economics](), [Forecasts](), [Technology](), [WebTech]() | [Permalink]()