Q: What is “big data,” and how does FRED-MD contribute to it?

A: Statistical analysis has evolved. In the past, it was focused on one variable measured across people or one variable measured across time. But with the advent of superfast computers, researchers and analysts can jointly model a large number of variables, each with a large number of observations across time. That is “big data.”

Although being able to use big data has benefits, such as improving the accuracy of forecasts, collecting the data can be extremely time-consuming. To that end, my co-author, Serena Ng of Columbia University, and I (along with tremendous assistance from staff at the St. Louis Fed’s data desk) created FRED-MD, a monthly database of over 130 macroeconomic time series that cover categories such as output and income, the labor market and prices. The data series are similar to the ones used by James Stock of Harvard and Mark Watson of Princeton, who created a macroeconomic data set that has become the benchmark for a lot of what people do in economics when they are working with big data.

With Stock and Watson’s choice of data as a guide, we used series that are available in FRED (Federal Reserve Economic Data), the St. Louis Fed’s main economic database. Now, rather than having thousands of economists separately put together their own data set, they can simply download a spreadsheet from our website.¹

FRED-MD has several advantages. For one, using series from FRED allows us to update our data set relatively quickly each month. In addition, anyone can access the latest file as well as previous vintages, which allows for easier replication of empirical work and for easier comparison between methods used in different lines of research. In other words, results won’t differ simply because the researchers used two different data sets. Another advantage of FRED-MD is that it saves users from having to incorporate revisions and changes to the data themselves. Those are handled by the experts at the data desk.

Our main goal in providing this core data set was to make it easier for those who do empirical analysis of big data. Instead of spending time collecting the data, they can focus on the bigger questions that they are trying to answer.

¹ For more information on FRED-MD and FRED-QD, which is a database of quarterly observations, see https://research.stlouisfed.org/econ/mccracken/fred-databases.