

## Big Data in economics

### New sources of data create challenges that may require new skills

Keywords: Big Data, machine learning, prediction, causal inference

#### ELEVATOR PITCH

Big Data refers to data sets of much larger size, higher frequency, and often more personalized information. Examples include data collected by smart sensors in homes or aggregation of tweets on Twitter. In small data sets, traditional econometric methods tend to outperform more complex techniques. In large data sets, however, machine learning methods shine. New analytic approaches are needed to make the most of Big Data in economics. Researchers and policymakers should thus pay close attention to recent developments in machine learning techniques if they want to fully take advantage of these new sources of Big Data.

#### KEY FINDINGS

##### Pros

- + Complex data are now available, characterized by large volume, fast velocity, diverse varieties, and the ability to link many data sets together.
- + Powerful new analytic techniques derived from machine learning are increasingly part of the mainstream econometric toolbox.
- + Big Data allows for better prediction of economic phenomena and improves causal inference.
- + Machine learning techniques allow researchers to create simple models that describe very large, complex data sets.
- + Machine learning methods and Big Data also allow for the complex modeling of relationships that predict well beyond the sample.

##### Cons

- Predictions based on Big Data may have privacy concerns.
- Machine learning methods are computationally intensive, may not have unique solutions, and may require a high degree of fine tuning for optimal performance.
- Big Data is costly to collect and store, and analyzing it requires investments in technology and human skill.
- Big Data may suffer from selection bias depending on how and by whom data are being generated.
- Access to these data may involve partnering with firms who limit researcher freedom.

#### AUTHOR'S MAIN MESSAGE

Due to the prevalence of connected digital devices, observational data sets are now available that are much larger and of higher frequency than traditional surveys: so-called Big Data. This has created opportunities for economists and policymakers to learn about economic systems and choices with a higher degree of precision. However, new methods, particularly those related to machine learning, are needed to take full advantage of Big Data. Furthermore, policymakers should consider a broader range of data as sensitive, researchers need checks to avoid unintentional bias, and economists should learn general-purpose coding languages.

