

Chapter 12

Big Data and Computational Social Science for Economic Analysis and Policy



Sebastiano Manzan

Abstract The goal of this chapter is to survey the recent applications of big data in economics and finance. An important advantage of these large alternative datasets is that they provide very detailed information about economic behaviour and decisions which has spurred research aiming at answering long-standing economic questions. Another relevant characteristic of these datasets is that they might be available in real time, a property that can be used to construct economic indicators at high frequencies. Overall, big alternative datasets have the potential to make an impact on economic research and policy and to complement the information used by governmental agencies to produce the official statistics.

12.1 Introduction

Computational social science (CSS) can be broadly defined as the area of the social sciences that makes computing power an essential tool to conduct the analysis. The field has a long tradition in economics that goes back to the 1970s when economists started to use computers to solve numerically economic models. Since then, there has been an exponential growth in applications as documented by four *Handbooks of Computational Economics* published between 1996 and 2018 (see Amman et al., 1996; Hommes & LeBaron, 2018; Schmedders & Judd, 2013; Tesfatsion & Judd, 2006). Computational economics can be broadly characterized in three main areas of activity: numerical methods to solve economic models, agent-based models, and computationally intensive techniques to analyse and model big datasets. The limited goal of this chapter is to provide an overview that focuses on the analysis and modelling of large datasets, while I refer to Fontana and Guerzoni (2023) in this Handbook for a review of agent-based models (ABM) and their use in economic problems. The availability of big data offers the possibility to investigate

S. Manzan (✉)

Zicklin School of Business, Baruch College, CUNY, New York, NY, USA

e-mail: sebastiano.manzan@baruch.cuny.edu

© The Author(s) 2023

E. Bertoni et al. (eds.), *Handbook of Computational Social Science for Policy*,

https://doi.org/10.1007/978-3-031-16624-2_12

long-lasting questions using more detailed information about economic behaviour. In addition, these datasets allow to uncover new empirical facts that were not previously known due to lack of information.

What exactly is “big data” in the context of economic applications? It can be defined as datasets that require advanced computing hardware and/or software tools to conduct the analysis. One such tool is distributed computing that shares the processing of a task across several machines, instead of a single machine as typically done by economists. Examples of large datasets used in economic analysis are administrative data (e.g. tax records for the whole population of a country), commercial datasets (e.g. consumer panels), and textual data (e.g. such as Twitter or news data) just to mention a few. In some cases, the datasets are structured and ready for analysis, while in other cases (e.g. text), the data is unstructured and requires a preliminary step to extract and organize the relevant information. As discussed in Einav and Levin (2014), economists are still in the early stages of analysing big data and are learning from developments in other disciplines. In particular, there is renewed interest in machine learning (ML) algorithms after the early applications of the 1990s (Kuan & White, 1994). Varian (2014) discusses techniques that can be used to analyse large datasets.

How can big data contribute to a better understanding of the economy and to support policy? In the highly aggregate context of macroeconomic analysis, big data offer the opportunity to bring to light the heterogeneity in consumers and firms that is typically neglected in official statistics. The high granularity of big datasets can be exploited to construct indicators that are better designed to explain certain phenomena, for example, along a geographic or demographic dimension. In addition, many economic models make assumptions about deep behavioural parameters that are difficult to estimate without detailed datasets. An example is represented by the work of Chetty et al. (2014b) where individual information about the school performance of a child is matched to his/her path of future earnings derived from tax data of the Internal Revenue Service (IRS). In other situations, big data allow to measure quantities that we could not measure until now. A field that is benefiting from these alternative sources of data is development economics. For instance, Storeygard (2016) uses night-light satellite data to estimate the income of sub-Saharan African cities.

Another important dimension in which big data can contribute to economic analysis is by offering information that is not only more granular but also more frequent in the time dimension. At times when economic conditions are rapidly changing, policy-makers need an accurate measure of the state of the economy to design the appropriate policy response. An example is provided by the early days of the Covid-19 pandemic in March 2020 when policy-makers felt the pressure to act in support of the economy despite the lack of official statistics to measure the extent of the slump, as discussed by Barbaglia et al. (2022). Many relevant economic indicators are observed infrequently, such as gross domestic product (GDP) at the quarterly frequency and the unemployment rate and the industrial production index at the monthly frequency. In addition, these variables are released with delays that range from a few days to several months. For these reasons, big data have the

potential to produce indicators of business conditions that are more accurate and timely.

More generally, private companies are amassing significant amounts of data that could be used to complement official statistics and inform economic policy. As discussed by Bostic et al. (2016), the approach of governmental agencies to produce official statistics is based, to a large extent, on consumer and business surveys. The approach guarantees the accuracy and the representativeness of the sample, although it comes at the cost of being an expensive and time-consuming exercise. Hence, the availability of alternative datasets offers the possibility of extracting information that can complement the evidence obtained from the surveys (a deeper analysis on the issue of the use of digital trace data and unconventional data in official statistics can be found in Signorelli et al., 2023).

However, we are also faced with a new set of issues regarding data governance and ethics issues as discussed by Taylor (2023) in this Handbook.

The chapter is organized as follows. I first review some of the recent work in economics and finance that leverages large datasets and emphasize the role of big data in allowing the researcher to conduct the analysis. I then draw some conclusions and discuss areas of potential development of the field.

12.2 Big Data in Economics

In this sect. 12.1, I discuss the main findings of recent applications of big data in economics and finance. I organize the discussion by *data source* with the intention to provide a more consistent review of the results. The goal of this section is not to be exhaustive, but rather to offer a concise overview of some of the main applications of big data to economics.

12.2.1 Administrative Data

Administrative data refer to data collected by governmental agencies as part of their mandate. As discussed by Card et al. (2010), the main advantages of administrative data, relative to surveys, are their large samples, the low attrition and non-response rates, and the small measurement error. In addition, administrative datasets are very detailed in terms of the information available regarding individuals. However, the researcher is confronted with significant challenges in conducting the analysis given the restricted access to the data. Typically, the researcher is required to provide the code to the government agency that actually conducts the analysis, slowing down significantly the development of the research project.

An influential paper using tax record data is Chetty et al. (2014). The goal of the paper is to investigate intergenerational mobility in the USA. They use a sample of 40 million children born between 1980 and 1982 and relate their income at age

30 to the parents' income. This administrative dataset represents a unique setting to evaluate intergenerational mobility since it provides a large sample going back to the 1980 and allows to link children and parents with very high accuracy.

Information from the Social Security Administration (SSA) is used in Kopczuk et al. (2010) to investigate income inequality and social mobility in the USA starting in 1937. They find that inequality decreased up to the early 1950s and increased steadily since then. In terms of social mobility, they show that it has been relatively constant over time, including at the top end of the income distribution.

Big administrative datasets are also used to evaluate educational attainments and teaching effectiveness. Dobbie and Fryer (2011) uses administrative data from the New York City Department of Education to evaluate the effect of charter school programmes on students' achievement. The evidence suggests that charter schools have a significant positive effect on improving the academic performance of poor children across several metrics. One of the possible explanations for these improvements is that the schools employ high-quality teachers. The issue of measuring the quality of the teachers and their impact on student performance is investigated by Chetty et al. (2014a) and Chetty et al. (2014b). They use a sample of one million students and match data from the school districts and tax records to track the evolution of earnings for the children in the sample. They find that measures of teacher's value added (VA), such as student's test scores, do not show a significant bias as proxies of teacher's quality. In addition, by matching students to their subsequent tax record, Chetty et al. (2014b) find that elementary school teachers with higher VA have a positive effect on college attendance and average earnings, among other measures.

Another source of administrative data is the credit register used by Jiménez et al. (2014) to evaluate the effect of monetary policy on bank's lending behaviour. The credit register records all loans and contracts between the public and the banking sector in a country. They show that a lower interest rate has the effect of increasing bank's risk-taking behaviour which leads to an increase in the supply of credit, in particular to more risky borrowers.

12.2.2 Financial Data

Financial transaction data represent a prominent source of big data in economic analysis. An early application is represented by Gross and Souleles (2002) that use a random sample of 24 thousand credit card accounts to investigate the effect on debt of changes to credit limits. Their results show that individuals respond to an increase in credit limits by borrowing more, in particular for those that started near the limit. Another more recent application using credit card data is Gallagher and Hartley (2017) that use a random 5% sample of individuals with credit history. They use hurricane Katrina as a natural experiment and find that households that lived in areas most affected by the flood experienced large reductions in debt, mostly due to the decline in home loan obligations. Horvath et al. (2021) use credit card data to

evaluate the behaviour of consumers during the 2020 pandemic. They find that credit card spending and balances declined rapidly during March/April 2020, in particular in areas with the highest incidence of cases. The recovery in spending started in May 2020 with riskier borrowers leading the way relative to those with high credit score. Dunn et al. (2020) use daily credit card data to assess the geographical and sectoral impact of the pandemic on consumer spending. They show that their measure of spending closely proxy for the monthly retail trade official statistic, which demonstrates the benefit of using big data to monitor the economy in real time. A similar analysis is provided in Bodas et al. (2019) and Carvalho et al. (2020) for Spain.

Calvet et al. (2009) use administrative data on the asset holdings and demographic information of all taxpayers in Sweden. The aim of the paper is to evaluate the financial sophistication of households in avoiding investment mistakes, such as under-diversification, inertia in risk taking, and holding losing stocks while selling winning stocks. They find that households with higher wealth and education levels are more sophisticated and less prone to investment mistakes.

12.2.3 Labour Markets

Labour market statistics have historically been data-rich due to the direct involvement of government agencies in the administration of unemployment benefits. Recently, private companies have started collecting information about the labour market. Naturally, the question is the representativeness of these private datasets for the overall labour market and the US economy. Horton and Tambe (2015) is a recent survey of the various sources of alternative labour market data that have emerged in recent years and provide a detailed discussion of the advantages and disadvantages of using such data. Napierala and Kvetan (2023) in this Handbook provide a complementary analysis of the role that big data can play in the analysis of the evolution of job skills.

An example of the use of alternative labour market data for policy is provided by Cajner et al. (2019). They use payroll data from the private company ADP to construct employment measures similar to those constructed by the Bureau of Labor Statistics (BLS) using the Current Employment Statistics (CES). They find that the two measures of employment complement each other and jointly they provide information about the dynamics of the labour market. This is a very important contribution since it shows that alternative data can provide information that is complementary and highly correlated with official statistics. The additional advantage of these private data sources is that they are available at higher frequencies and allow the researcher to segment the sample geographically and by demographic characteristics. This benefit is discussed in Cajner et al. (2020) that shows the real-time behaviour of the weekly employment measure during the Covid-19 pandemic relative to the monthly official statistic from CES. Similar results are also obtained by Gregory and Zhu (2014).

12.2.4 Textual Data

An alternative source of data that is gaining interest in economics and finance is textual data. In this case, the goal is to use text from newspapers, speeches, company reports, and Twitter, among others, to construct measures that help understand economic behaviour or predict economic variables. Gentzkow et al. (2019) provide a recent overview of the work done so far.

An important source of text data is newspaper articles that might be considered a proxy for the information set available to the public when making an economic decision. An early paper is Tetlock (2007) that extracts sentiment from a column of *The Wall Street Journal* and finds that it is useful to predict daily returns of the aggregate market. Baker et al. (2016) aim at measuring economic and political uncertainty by counting the number of articles that contain a set of keywords associated with uncertainty. They show that their measure is highly correlated with measures of uncertainty. Other recent applications analyse news to construct proxies for economic sentiment (see Barbaglia et al., forthcoming; Larsen & Thorsrud, 2019; Shapiro et al., 2020; Thorsrud, 2020). Monitoring the sentiment of consumers and businesses has a long tradition in economics, and it is typically based on surveys. The contribution of these papers is to show that sentiment based on newspaper articles has a similar behaviour to survey-based sentiment. These indicators are found to have forecasting power for several macroeconomic variables that is incremental relative to the typical macroeconomic predictors (Barbaglia et al., forthcoming). Larsen and Thorsrud (2019) investigate the relation between news and consumer expectations and find that the topics extracted from the news contribute to explain the consumers' decision to update their inflation expectations.

Another line of research has investigated the role of communication in the implementation of monetary policy. Hansen and McMahon (2016) use the text of verbal and written communication by the Federal Reserve to understand its role in predicting economic variables. They find that the forward guidance embedded in the central bank statements is more relevant relative to the communication of the state of the economy. Hansen et al. (2018) investigate the role of increasing transparency in the central bank communication by analysing the internal deliberation of the policy-makers. They find that their communication patterns changed significantly after transparency was introduced.

The GDELT project¹ is another source of textual data that has been used in several applications. Consoli et al. (2021) use sentiment analysis to understand the dynamics of sovereign yields in Europe. Acemoglu et al. (2018) use GDELT to identify events of political and social unrest in Egypt and to evaluate their effect on stock returns.

A data source that is gathering momentum in economic and financial analysis is Twitter. Baker et al. (2021) use Twitter messages to construct a Twitter Economic

¹ More information about GDELT is available at <https://www.gdelproject.org/>.

Uncertainty (TEU) indicator similar to the EPU indicator proposed by Baker et al. (2021) that is based on newspaper articles. Their results show that there is a very high correlation between TEU and EPU.

12.2.5 Mobile Phone Data

Mobile phone data represents an additional source of big data for economic analysis. This type of data is potentially very high dimensional since it tracks the location of a user over time. An economic application is represented by Blumenstock et al. (2015) that use mobile phone data to measure the socio-economic status of the caller. This is a particularly useful initiative for developing countries where official statistics are not very reliable and well-developed. Milusheva (2020) uses mobile phone data to track the effect of the movement of people from high-disease areas to low-disease areas on malaria spreading. A similar idea is developed in Iacus et al. (2020) that investigate the effect of the containment measures on the spreading of the Covid-19 virus. Their findings suggest that a measure of mobility constructed from mobile phone data is a highly accurate predictor of the initial spread of the virus in Italy and France.

12.2.6 Internet Data

The emergence of the internet has created the opportunity for researchers to collect online data to proxy for economic variables of interest (see Edelman, 2012, for a detailed discussion). An example is provided by the emergence of eBay as a marketplace for the exchange of goods that allowed economists to test market design mechanisms and to investigate the behaviour of bidders and sellers. An early paper is Bajari and Hortacsu (2003) that examine the empirical regularities of eBay auctions and estimate a model of bidding.

An area of intense recent work has been measuring social ties based on online platform, such as Facebook. Bailey et al. (2018a) discuss the construction of the Social Connectedness Index (SCI) which measures the friendship connections between Facebook users living in different geographical areas of the USA and abroad. An application of the SCI to explain the housing market is provided in Bailey et al. (2018b). They find that social connections contribute to explain the surge in house prices which they argue to be the result of the similarity of experience and expectations about the housing market.

Cavallo and Rigobon (2016) uses price data that are scraped from online stores to construct measures of inflation. These measures are found to track well the official statistics and have the advantage that can be calculated at high frequencies. Goolsbee and Klenow (2018) use a large dataset of e-commerce transactions to

calculate the inflation rate. They find that during the period 2014–2017, the inflation rate was 3% lower relative to the official Consumer Price Index (CPI).

Another big dataset that has recently gained interest among economists is Google Trends. It represents a measure of the intensity of queries in the Google search engine regarding a set of keywords in a certain geographic area. The big data feature of Google Trends is that the time series for the search terms is the outcome of the aggregation across millions of queries by Google users around the world. Google Trends can be interpreted as a sentiment measure since it captures the public interest on a specific topic at a certain point in time. An early contribution using Google Trends is Choi and Varian (2012) that finds that including appropriately selected trends improves the accuracy of nowcasts for several economic variables. D'Amuri and Marcucci (2017) use job search-related queries to forecast the unemployment rate in the USA. Their results show that using Google Trends improves accuracy also relative to professional forecasters and are particularly accurate during turning points that are difficult to predict in real time. Castelnuovo and Tran (2017) construct an indicator that they call Google Trends Uncertainty (GTU) that aims at capturing Economic and Political Uncertainty (EPU) in the spirit of Baker et al. (2016) using series from Google Trends.

12.2.7 Other Data

An interesting application of seismic data to economics is represented by Tiozzo Pezzoli and Tosetti (2021). They use seismic data to identify vibrations produced by human activity, such as air and road traffic and manufacturing activity among others. They find that the indicator they construct is strongly correlated with several official measures of economic activity.

Another source of alternative big data is obtained from satellite images that are used in a variety of CSS applications. However, only recently, economists realized the potential of satellite image data for economic analysis. Donaldson and Storeygard (2016) and Gibson et al. (2020) provide overviews of the application of satellite data in economics and a primer on remote sensing.

Chen and Nordhaus (2011) use night-light satellite data to improve GDP measures for developing countries, which is particularly relevant when official statistics are missing. The paper shows that luminosity provides informational value that can help improve the accuracy of output measures. Galimberti (2020) performs a similar exercise with the focus on the forecasting ability of the measures of economic activity based on the luminosity data. The results indicate that these measures are useful to improve the accuracy of simple forecasting models, although country-specific models deliver better forecast performance relative to the pooled model. In a similar context, Hu and Yao (2021) propose an econometric methodology to use luminosity to improve GDP measures. Henderson et al. (2011) provides a detailed discussion of applications of night lights to measure national income, in particular in the case of developing economies. Another application of night-light

data is represented by Storeygard (2016) that evaluates whether the distance of cities from a port influenced their growth in sub-Saharan African countries. The role of the satellite data in this case is to provide a measure of economic activity at the city level that are not otherwise available from official statistics.

12.3 Conclusion

The discussion in this chapter demonstrates how big data can be valuable to answer long-standing questions and to test the validity of economic assumptions. An illustration is the work with administrative data discussed earlier that shows the great potential of providing economic researchers access to these data, but highlights also the severe limitations of scaling up the availability of these data to a wider audience of users. Another challenge is represented by the fact that many of these alternative datasets are collected by private companies that might have low incentives to share the data with researchers. However, big data have a significant public role to play which calls for a framework that facilitates sharing of the information. An example of the public relevance of using big data is to produce real-time indicators of business conditions. In this respect, the collaboration between the Federal Reserve and the payroll processor ADP (Cajner et al., 2019) indicates how the private big dataset can complement the existing information provided by statistical agencies to support economic policy in real time. This collaboration is likely to set the path for more extensive partnerships between the private sector and statistical agencies. As argued in Bostic et al. (2016), the current model of the production of economic data is the domain of governmental agencies that are funding and running the collection of data, typically in the form of consumer and business surveys. This model is likely to evolve in the future as companies collect increasing amounts of economic data that are valuable, and most likely cheaper, to the production of official statistics.

Acknowledgments The author is grateful to Eleonora Bertoni, Matteo Fontana, Lorenzo Gabrielli, Serena Signorelli, Michele Vespe, and Luca Barbaglia of the Joint Research Centre of the European Commission for the helpful comments that have improved the organization and clarity of the chapter.

References

- Acemoglu, D., Hassan, T. A., & Tahoun, A. (2018). The power of the street: Evidence from Egypt's Arab Spring. *Review of Financial Studies*, 31(1), 1–42.
- Amman, H. M., Tesfatsion, L., Kendrick, D. A., Rust, J., Judd, K. L., Schmedders, K., Hommes, C. H., & LeBaron, B. D. (1996). *Handbook of computational economics: Agent-based computational economics* (Vol. 2). Elsevier.
- Bailey, M., Cao, R., Kuchler, T., Stroebel, J., & Wong, A. (2018a). Social connectedness: Measurement, determinants, and effects. *Journal of Economic Perspectives*, 32(3), 259–80.

- Bailey, M., Cao, R., Kuchler, T., & Stroebel, J. (2018b). The economic effects of social networks: Evidence from the housing market. *Journal of Political Economy*, 126(6), 2224–2276.
- Bajari, P., & Hortacsu, A. (2003). The winner's curse, reserve prices, and endogenous entry: Empirical insights from ebay auctions. *Rand Journal of Economics*, 34, 329–355.
- Baker, S. R., Bloom, N., Davis, S., & Renault, T. (2021). Twitter-derived measures of economic uncertainty. *Working paper*.
- Baker, S. R., Bloom, N., & Davis, S. J. (2016). Measuring economic policy uncertainty. *Quarterly Journal of Economics*, 131(4), 1593–1636.
- Barbaglia, L., Consoli, S., & Manzan, S. (forthcoming). Forecasting with economic news. *Journal of Business and Economic Statistics*.
- Barbaglia, L., Frattarolo, L., Onorante, L., Pericoli, F. M., Ratto, M., & Pezzoli, L. T. (2022). Testing big data in a big crisis: Nowcasting under COVID-19. *International Journal of Forecasting*, <https://doi.org/10.1016/j.ijforecast.2022.10.005>
- Blumenstock, J., Cadamuro, G., & On, R. (2015). Predicting poverty and wealth from mobile phone metadata. *Science*, 350(6264), 1073–1076.
- Bodas, D., Garcia Lopez, J. R., Murillo Arias, J., Pacce, M. J., Rodrigo López, T., Romero Palop, J. d. D., Ruiz de Aguirre, P., Ulloa Ariza, C. A., & Valero Lapaz, H. (2019). Measuring retail trade using card transactional data. *Documentos de trabajo/Banco de España*, 1921.
- Bostic, W. G., Jarmin, R. S., & Moyer, B. (2016). Modernizing federal economic statistics. *American Economic Review*, 106(5), 161–64.
- Cajner, T., Crane, L., Decker, R., Hamins-Puertolas, A., Kurz, C. J., et al. (2019). Tracking the labor market with “Big Data”. *Working paper*.
- Cajner, T., Crane, L. D., Decker, R., Hamins-Puertolas, A., & Kurz, C. J. (2020). Tracking labor market developments during the Covid-19 pandemic: A preliminary assessment. *Working paper*.
- Calvet, L. E., Campbell, J. Y., & Sodini, P. (2009). Measuring the financial sophistication of households. *American Economic Review*, 99(2), 393–98.
- Card, D., Chetty, R., Feldstein, M. S., & Saez, E. (2010). Expanding access to administrative data for research in the United States. *Working paper*.
- Carvalho, V. M., Hansen, S., Ortiz, A., Garcia, J. R., Rodrigo, T., Rodríguez Mora, S., & Ruiz de Aguirre, P. (2020). Tracking the COVID-19 crisis with high-resolution transaction data. *Working paper*.
- Castelnuovo, E., & Tran, T. D. (2017). Google it up! a Google Trends-based uncertainty index for the United States and Australia. *Economics Letters*, 161, 149–153.
- Cavallo, A., & Rigobon, R. (2016). The billion prices project: Using online prices for measurement and research. *Journal of Economic Perspectives*, 30(2), 151–78.
- Chen, X., & Nordhaus, W. D. (2011). Using luminosity data as a proxy for economic statistics. *Proceedings of the National Academy of Sciences*, 108(21), 8589–8594.
- Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014a). Measuring the impacts of teachers I: Evaluating bias in teacher value-added estimates. *American Economic Review*, 104(9), 2593–2632.
- Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014b). Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood. *American Economic Review*, 104(9), 2633–79.
- Chetty, R., Hendren, N., Kline, P., & Saez, E. (2014). Where is the land of opportunity? the geography of intergenerational mobility in the United States. *Quarterly Journal of Economics*, 129(4), 1553–1623.
- Choi, H., & Varian, H. (2012). Predicting the present with Google Trends. *Economic Record*, 88, 2–9.
- Consoli, S., Pezzoli, L. T., & Tosetti, E. (2021). Emotions in macroeconomic news and their impact on the european bond market. *Journal of International Money and Finance*, 118, 102472.
- D'Amuri, F., & Marcucci, J. (2017). The predictive power of Google searches in forecasting US unemployment. *International Journal of Forecasting*, 33(4), 801–816.

- Dobbie, W., & Fryer, R. G. (2011). Are high-quality schools enough to increase achievement among the poor? Evidence from the Harlem Children's Zone. *American Economic Journal: Applied Economics*, 3(3), 158–87.
- Donaldson, D., & Storeygard, A. (2016). The view from above: Applications of satellite data in economics. *Journal of Economic Perspectives*, 30(4), 171–98.
- Dunn, A., Hood, K., & Driessen, A. (2020). Measuring the effects of the COVID-19 pandemic on consumer spending using card transaction data. *Working paper*.
- Edelman, B. (2012). Using internet data for economic research. *Journal of Economic Perspectives*, 26(2), 189–206.
- Einav, L., & Levin, J. (2014). Economics in the age of big data. *Science*, 346(6210), 1243089.
- Fontana, M., & Guerzoni, M. (2023). Modeling complexity with unconventional data: Foundational issues in computational social science. In Bertoni, E., Fontana, M., Gabrielli, L., Signorelli, S., & Vespe, M. (Eds.), *Handbook of computational social science for policy*. Springer.
- Galimberti, J. K. (2020). Forecasting GDP growth from outer space. *Oxford Bulletin of Economics and Statistics*, 82(4), 697–722.
- Gallagher, J., & Hartley, D. (2017). Household finance after a natural disaster: The case of hurricane Katrina. *American Economic Journal: Economic Policy*, 9(3), 199–228.
- Gentzkow, M., Kelly, B., & Taddy, M. (2019). Text as data. *Journal of Economic Literature*, 57(3), 535–74.
- Gibson, J., Olivia, S., & Boe-Gibson, G. (2020). Night lights in economics: Sources and uses 1. *Journal of Economic Surveys*, 34(5), 955–980.
- Goolsbee, A. D., & Klenow, P. J. (2018). Internet rising, prices falling: Measuring inflation in a world of e-commerce. *Aea Papers and Proceedings*, 108, 488–92.
- Gregory, A. W., & Zhu, H. (2014). Testing the value of lead information in forecasting monthly changes in employment from the bureau of labor statistics. *Applied Financial Economics*, 24(7), 505–514.
- Gross, D. B., & Souleles, N. S. (2002). Do liquidity constraints and interest rates matter for consumer behavior? evidence from credit card data. *Quarterly Journal of Economics*, 117(1), 149–185.
- Hansen, S., & McMahon, M. (2016). Shocking language: Understanding the macroeconomic effects of central bank communication. *Journal of International Economics*, 99, S114–S133.
- Hansen, S., McMahon, M., & Prat, A. (2018). Transparency and deliberation within the FOMC: A computational linguistics approach. *The Quarterly Journal of Economics*, 133(2), 801–870.
- Henderson, V., Storeygard, A., & Weil, D. N. (2011). A bright idea for measuring economic growth. *American Economic Review*, 101(3), 194–99.
- Hommes, C., & LeBaron, B. (2018). *Computational economics: Heterogeneous agent modeling*. Elsevier.
- Horton, J. J., & Tambe, P. (2015). Labor economists get their microscope: Big data and labor market analysis. *Big Data*, 3(3), 130–137.
- Horvath, A., Kay, B. S., & Wix, C. (2021). The Covid-19 shock and consumer credit: Evidence from credit card data. *Working paper*.
- Hu, Y., & Yao, J. (2021). Illuminating economic growth. *Journal of Econometrics*, 228(2), 359–378.
- Iacus, S. M., Santamaria, C., Sermi, F., Spyros, S., Tarchi, D., & Vespe, M. (2020). Human mobility and covid-19 initial dynamics. *Nonlinear Dynamics*, 101(3), 1901–1919.
- Jiménez, G., Ongena, S., Peydró, J.-L., & Saurina, J. (2014). Hazardous times for monetary policy: What do twenty-three million bank loans say about the effects of monetary policy on credit risk-taking? *Econometrica*, 82(2), 463–505.
- Kopczuk, W., Saez, E., & Song, J. (2010). Earnings inequality and mobility in the United States: Evidence from social security data since 1937. *Quarterly Journal of Economics*, 125(1), 91–128.
- Kuan, C.-M., & White, H. (1994). Artificial neural networks: An econometric perspective. *Econometric Reviews*, 13(1), 1–91.

- Larsen, V. H., & Thorsrud, L. A. (2019). The value of news for economic developments. *Journal of Econometrics*, 210(1), 203–218.
- Milusheva, S. (2020). Managing the spread of disease with mobile phone data. *Journal of Development Economics*, 147, 102559.
- Napierala, J., & Kvetan, V. (2023). Changing job skills in a changing world. In Bertoni, E., Fontana, M., Gabrielli, L., Signorelli, S., & Vespe, M. (Eds.), *Handbook of computational social science*. Springer.
- Schmedders, K., & Judd, K. L. (2013). *Handbook of computational economics*. Newnes.
- Shapiro, A. H., Sudhof, M., & Wilson, D. J. (2020). Measuring news sentiment. *Journal of Econometrics*, 228(2), 221–243.
- Signorelli, S., Fontana, M., Gabrielli, L., & Vespe, M. (2023). Challenges for official statistics in the digital age. In Bertoni, E., Fontana, M., Gabrielli, L., Signorelli, S., & Vespe, M. (Eds.), *Handbook of computational social science for policy*. Springer.
- Storeygard, A. (2016). Farther on down the road: Transport costs, trade and urban growth in Sub-Saharan Africa. *Review of Economic Studies*, 83(3), 1263–1295.
- Taylor, L. (2023). Data justice, computational social science and policy. Bertoni, E., Fontana, M., Gabrielli, L., Signorelli, S., & Vespe, M. (Eds.), *Handbook of computational social science*. Springer.
- Tesfatsion, L., & Judd, K. L. (2006). *Handbook of computational economics: Agent-based computational economics*. Elsevier.
- Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. *Journal of Finance*, 62(3), 1139–1168.
- Thorsrud, L. A. (2020). Words are the new numbers: A newsy coincident index of the business cycle. *Journal of Business & Economic Statistics*, 38(2), 393–409.
- Tiozzo Pezzoli, L., & Tosetti, E. (2021). Seismonomics: Listening to the heartbeat of the economy. *Working paper*.
- Varian, H. R. (2014). Big data: New tricks for econometrics. *Journal of Economic Perspectives*, 28(2), 3–28.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

