DATA, DIGITIZATION AND PRODUCTIVITY

C. CORRADO, J. HASKEL, M. IOMMI. C. JONA-LASINIO

JUNE 13TH 2021.

CONTACT: CORRADO, CAROL <CAROL.CORRADO@CONFERENCE-BOARD.ORG>

I. DATA IN AN EXPANDED INVESTMENT FRAMEWORK: CONCEPTS

With growing data and digitization, there is a flowering of work on the importance of data in understanding modern economies. Much of that the work focuses on "bigdata" and develops implications for business strategy and "the data value chain". By a very long way, the best-selling book on this subject is (Mayer-Schonberger and Cukier 2013) "Big Data: A Revolution That Will Transform How We Live, Work, and Think". That book makes a number of statistical claims, such as suggesting that Big Data will be used to disprove much casually held causal intuition and can reduce measurement problems. It also uses a conceptual framework, classifying the properties of bigdata according to the three Vs (volume, variety, and velocity) and, using a rich variety of case studies, documents how companies have managed to profit from data in various ways. In the meantime, The Economist (2017) suggested "data is the new oil" and Google's Eric Schmidt is commonly quoted as stating that as much data/information is being created every two days as was created from the dawn of civilization to 2003 (Wong 2012).

The technology literature commonly refers to the "data stack", describing the transformation of raw records to usable data and insights, somewhat analogous to the "data value chain" in the management/business strategy literature.

Meantime, it is also worth mentioning the proliferation of terms used almost interchangeably, at least in the United Kingdom. Official definitions state that the *Data* Protection Act "controls how your *personal information* is used," that the UK Information Commissioner "promotes *data* privacy for individuals", and that the Freedom of *Information* Act allows citizens to request publicly held *datasets*.

To put this into a framework amenable for measurement and focusing on the economic properties of data, rather than the describing its features (e.g. Big or Small amounts of it), we make the case for treating data, and the knowledge gleaned from data analytics, as long-lived assets that contribute to final production. To achieve this, firms must *invest*: both in processing raw data into analytical insights (the data stack) and then combining such insights with other intangible and tangible assets to commercialize such insights (roughly the information value chain).

Thus "data" includes "bigdata" as used in the business literature, or even "massive data," a term frequently used in scientific applications. Bigdata has been defined in a technical sense by, e.g., the SAS Institute, as data that is so large, fast, or complex that it is difficult or impossible to process using traditional methods.

In an economic sense, data is viewed as an asset. Such an asset produces a flow of services. Such services are nonrival, yet likely to be excludable and then a factor input to production in modern economies. Conventional data processing long has been one of the primary functions of computing; indeed, at the dawn of computing technology (in the 1950s), modern industrial economies already contained significant information-processing activity (Machlup 1962). Data (i.e., bigdata) as analyzed in this paper is then

inextricably linked with computing and related technologies, manifest in the form of machine learning algorithms and other data analytic and artificial intelligence (AI) tools. Advances in AI have been especially rapid of late, e.g., machines became better than humans at image recognition circa 2015.

We define three types of data:

- *Raw records* are data stores not yet cleaned, formatted, or transformed ready for analysis, e.g., data scraped from the web, transactions, or sensor data. Raw records also cover the raw data collected from experiments, statistical surveys, or administrative records.
- *Databases* consist of transformed raw data, records that have been cleaned, formatted, combined and/or structured such that they are suitable for some form of data analytics or visualizations.
- *Data intelligence* reflects the further integration of data with advanced analytic tools; data intelligence (or data-derived knowledge, e.g., a new marketing strategy) is a set of quantitative inputs that provide actionable guidance for business decision-making (or solutions to scientific problems).

These categories align with categories set out in a conceptual framework for measuring data at Statistics Canada.

We will then look at *data monetization* which we define as occurring upon implementation of the actionable guidance yielded by data intelligence. Implementation uses the new intelligence, along with labor and capital (tangible and intangible) to deliver final goods and services. As with intangible assets, partial appropriability provides the basis for treating long-lived business spending to develop data as capital spending and to include data assets as productive capital inputs in productivity analysis. Investment in data readily fits into the expanded investment framework used for intangible capital due to (Corrado, Hulten, and Sichel 2005, 2009).

II. DATA IN AN EXPANDED INVESTMENT FRAMEWORK: measurement

Regarding national accounts, *data stores* are generally missing in our national statistics although an expansion of the asset boundary to include them in the next edition of the SNA is under discussion. *Databases* are included in the national accounts asset boundary (blended with software, at least in principle). *Data intelligence* is not included, except for its possible overlap with software and R&D.

Regarding the expanded investment framework used for intangible capital, both *databases* and *data intelligence* are notionally included, but they are not separately identified asset categories. A key challenge for the empirical analysis of data and data-driven economic activity is to identify the overlap between data assets contained, but not separately identified, in the Corrado, Hulten, Sichel (CHS) intangible asset types. We shall review this and rely on the analysis in Corrado, Haskel, Iommi, and Jona-Lasinio (2019), (Statistics Canada 2019a, 2019b) and reviews by (a) (Edquist, Goodridge, and Haskel 2021), who focused on the coverage of data and data science in the national accounts categories of software and R&D in the UK and selected countries in Europe, and (b) Corrado (2020), who looked for investments in AI and data analytics in R&D and other CHS intangibles in the US.

Our preliminary conclusions are that a fraction of (perhaps one half) investments in *data, databases, and AI application tools* likely is included in the current software investment component of GDP (though in some countries, e.g., the UK, the coverage likely is higher). Investments in the *development of new AI and data science tools and data intelligence* appear to be included in R&D in some countries (e.g., in the US),

whereas coverage of data-related R&D seems lacking in others (e.g., Canada, UK, and possibly much of Europe). Finally, although expenditures on data analytics is not a separately identified category, inspection of unpublished detail suggests the size of such spending is significant, i.e., it likely comprised at least one-third of the other CHS asset categories in the United States in 2018, having grown steadily in relative terms since 2005.

We will also comment on the evidence on prices and depreciation of data. Likewise, we review what we know about the economic properties of data. This is relevant since an important stream of the macro literature treats data as showing economies of scale and excludability and thus are a source of market power for firms in many macro models. Interestingly, the micro evidence suggests (perhaps counter-intuitively) there are not economies of scale in data (Goldfarb and Tucker (2019)) .

III. DATA IN AN EXPANDED INVESTMENT FRAMEWORK: productivity

The increased use of data is commonly thought to be one of the drivers of the change in knowledge economies, especially in the aftermath of the global financial crisis (GFC). Even though new digital technologies have proliferated in recent years, global productivity growth has slowed. The confounding trends in data/AI use and measured productivity growth create a paradox of its own, suggesting the need for a thorough examination of the channels through which data use might have aggregate consequences (Corrado et al 2021). We use these insights to see what impact the trend to data and it's changed use has had on productivity.

We are in the process of creating a new edition of EUKLEMS combined with INTAN-Invest, i.e., a harmonized dataset including all Corrado-Hulten-Sichel intangibles covering Europe, the United States and Japan. (The four of us have a grant from the European Commission to construct this new intangibles-extended macro/industry dataset and preliminary results will be in time for preparing a paper for the conference). We shall present new cross-country results from this dataset and how the inclusion of data matters for productivity growth, depending, essentially on how much data is currently in software and other intangible assets.

References

Corrado, Carol, Charles Hulten, and Daniel Sichel. 2005. "Measuring Capital and Technology: An Expanded Framework." In *Measuring Capital in the New Economy*, edited by Carol Corrado, John Haltiwanger, and Daniel Sichel, 11–46. Studies in Income and Wealth No. 65. Chicago: University of Chicago Press.

———. 2009. "Intangible Capital And U.S. Economic Growth." *Review of Income and Wealth* 55 (3): 661–85. https://doi.org/10.1111/j.1475-4991.2009.00343.x.

Corrado C., J. Haskel, C. Jona-Lasinio, M.Iommi (2019), "Intangible Capital, Innovation, and Productivity *à la* Jorgenson", in Fraumeni, Barbara M. (ed.) *Measuring Economic Growth and Productivity*: *Foundations, KLEMS Production Models, and Extensions*, Academic Press, Cambridge, MA. **ISBN:** 9780128175965, 15 November 2019.

___(2021) The Value of Data in Digital-based Business Models: Measurement and Economic Policy Implications, forthcoming OECD report.

Edquist, Harald, P Goodridge, and J Haskel. 2021. "We See Data Everywhere except in the Productivity Statistics."

https://www.researchgate.net/publication/346416311_We_see_data_everywhere_except_in_the_productivity_statistics.

Goldfarb, Avi, and Catherine Tucker. 2019. "Digital Economics." *Journal of Economic Literature*, 57 (1): 3-43.DOI: 10.1257/jel.20171452

Machlup, Fritz. 1962. *The Production and Distribution of Knowledge in the United States*. Princeton, New Jersey: Princeton: Princeton University Press.

Mayer-Schonberger, Viktor, and Kenneth Cukier. 2013. *Big Data: A Revolution That Will Transform How We Live, Work, and Think*.

Statistics Canada. 2019a. “Measuring Investment in Data, Databases and Data Science: Conceptual Framework.” 2019. https://www150.statcan.gc.ca/n1/pub/13-605-x/2019001/article/00008-eng.htm.

———. 2019b. “The Value of Data in Canada: Experimental Estimates.” 2019. https://www150.statcan.gc.ca/n1/pub/13-605-x/2019001/article/00009-eng.htm.