



Valuing the U.S. Data Economy Using Machine Learning and Online Job Postings

Discussant: Carol Robbins

Date: May 13, 2022

Location: BEA Advisory Committee

This paper represents the views of the presenter and not necessarily those of NCSES.

National Center for Science and Engineering Statistics

Social, Behavioral and Economic Sciences

National Science Foundation

Contributions of this paper

- Address a critical data gap: own-account data bases
- Incorporates administrative data (online job advertisements) and machine learning to identify occupations producing data assets
- Improve input measures for cost accounting
 - Time-use factors
 - Burning Glass Technologies data
 - BLS Occupation/Industry data
- A big step forward for input-cost measurement

Uses of Intangible Investment Measures

- Performance assessment and benchmarking
- Informing public policy decisions
- Informing private sector policy decisions
- Facilitating social science research
- Growth accounting framework
 - Physical Capital, L= Labor Input, K = Knowledge Capital

Reference: Bronwyn H. Hall and Adam B. Jaffe (2018), "Measuring Science, Technology, and Innovation: A Review", *Annals of Science and Technology Policy*: Vol. 2: No. 1, pp 1-74. <http://dx.doi.org/10.1561/110.00000005>

Measuring Intangibles (Intellectual Property Products)

Nominal Expenditures (when available)

- Prepackaged and Custom Software
- Entertainment, Artistic, and Literary Originals
- Prepackaged and Custom databases

Cost-based measures (second best)

- R&D based on survey data on expenditures from NSF and Census Surveys
- Own-account software based on a fraction of time for pre-selected occupations

Which Workers Create own-account Databases?

- Identify skills unique to databases
 - Burning Glass Technologies 17,000 skills
 - Lancaster, et al 2021
 - 203 skills data relevant
 - 20 “landmark” skills (Blackburn, 2021)
- Link to occupations
 - O*Net SOC 2010
 - Estimate share of time based on skills overlaps
- Link to industries
 - BLS occupation by industry data
- Estimate overlaps
 - Survey-based R&D
 - 50% adjustment for R&D intensive industries

Highlights in Accounting for Intangibles

Measurement Progress	BEA Comprehensive Revision Year
Software	1999
R&D	2013
Entertainment, Literary, and Artistic Originals	2013
Improved own-account software and R&D estimates	2018
Data Assets	?

Authors fill a measurement gap in digital assets

- Own-account data assets
 - \$82.6 billion in 2003
 - \$159.5 billion 2020
 - Annual growth rate almost 4%
- Questions for discussion
 - Can the double-counting adjustments be improved?
 - For productivity analysis, which assets are freely available?
 - What other intangibles could be analyzed this way?



Contact: crobbins@nsf.gov

Context: Labor Input for Own-account Software

Person-months of Effort

Two fixed parameters: development effort and development time

- Effort is a nonlinear function of complexity and kilo lines of code (KLoC)

$$Effort = 2.4 (KLoC)^{1.05}$$

- Development time is a nonlinear function of effort

$$Nominal\ development\ time = 2.5 (Effort)^{0.38}$$

Note: 1.05 and 0.38 are COCOMO II parameters for a well understood application program, where the size of the development team is reasonably small, and the team members are experienced in developing similar types of projects. (B. W. Boehm, R. Madachy, 2000)