Valuing Stocks and Flows of Data Assets for the U.S. Business Sector

José Bayoán Santiago Calderón
Dylan G. Rassier

BEA Advisory Committee Meeting
May 13, 2022

Outline

- Background on Data as an Asset
- Methodology
- Experimental Results
- Conclusions and Next Steps

Outline

- Background on Data as an Asset
- Methodology
- Experimental Results
- Conclusions and Next Steps

Data Uses in Production: IoT Sensor Data

- Caterpillar and John Deere install sensors on machinery and equipment to collect data on operations.
- Sensor data are combined with historical and real-time data on weather, irrigation, and other relevant systems.
- Actionable intelligence is available in platforms to reduce downtime, save fuel, improve safety, and manage crops and worksites.

Data Uses in Production: Personal Data

- Insurance firms collect health and lifestyle data from fitness devices and social network sites to assess risk.
- Location firms collect geolocation data from mobile service providers and mobile apps to sell to advertisers, retailers, and investment firms.
- E-commerce firms collect purchase and browsing data to predict demand and improve service.

Literature on Data as an Asset

- Personal privacy
  - Acquisti, Taylor, Wagman (2016, *JEL*)
- Implications for productivity and economic growth
  - Farboodi, Veldkamp (2021, NBER WP)
  - Jones, Tonetti (2020, *AER*)
  - Goodridge, Haskel, Edquist (2021, *ROIW*)
- Value of data
  - Varian (2018, NBER WP)
  - Hughes-Cromwick, Coronado (2019, *JEP*)
- How to measure and record data in national accounts?

System of National Accounts 2008 (SNA)

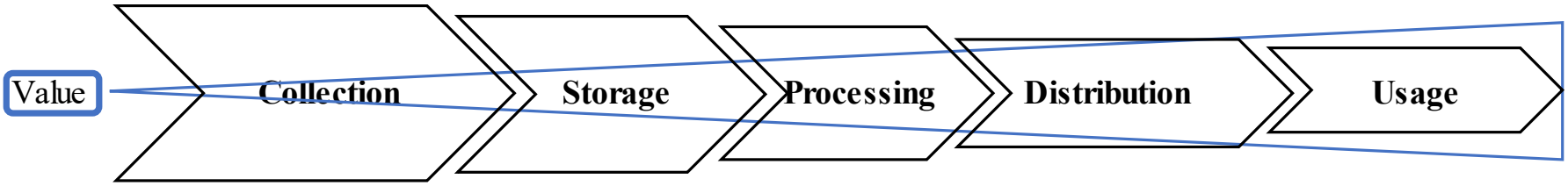
- Computer software and databases (*SNA* ¶10.109-114)
  - Category of intellectual property products (IPPs)
  - Grouped together in practice

- Definition of databases
  - Files of data organized in such a way as to permit resource-effective access and use of data.
    - Value should include the cost of preparing data in the appropriate format but exclude the cost of acquiring or producing data.
    - Databases for sale should be valued at their market price, which includes the value of the information content.
  - Inconsistent treatment for own-account and purchased data
    - Exclude value of data in own-account databases
    - Include value of data in market purchases of databases

System of National Accounts 2008 (SNA)

- Sentiment during the last revision of the *SNA*
  - Data is a knowledge asset that is not produced
  - Value of data is reflected in purchased goodwill

- Is data produced or non-produced?



Collection	Storage	Processing	Distribution	Usage
Electronic payments	Cloud facilities	Analysis	Tracking companies	Business models
Web-connected devices	Data centers	Fusion	Data exchanges	Regulatory programs
Regulatory filings	Servers		Data brokers	Social programs
Surveys			Ad networks	Information
				Entertainment

Source: Adapted from OECD (2013) and Visconti et al. (2017).

SNA Digitalization Task Team (DZTT)

- Data is a produced asset
  - Information content that is produced by *accessing* and *observing* phenomena and *recording*, *storing*, and *organizing* information elements from the phenomena in a digital format, which provide economic benefits in production.

- Expand the production boundary (i.e., GDP ↑)
  - Subcategory of computer software and databases
  - Sum of costs = labor + capital + intermediate inputs
  - Why and why now? More complete picture of intangibles used in production in response to data proliferation

- Forthcoming guidance in *SNA 2025*

IPPs in the U.S. National Accounts, 2020

Total investment.....*$1,078.5*

Software.....453.4

- Prepackaged.....$212.5
- Custom.....165.5
- Own-account.....75.4

Research and development.....537.7

Entertainment, literary, and artistic originals.....87.4

Mineral exploration and evaluation.....N/A

Other IPPs.....N/A

Source: NIPA Table 5.6.5. Billions USD. Mineral exploration costs are included in NIPA non-residential structures.

Measurement Challenges for IPPs

- Measuring capital costs
  - Does capital play a unique role for data as an asset?
- Multiple counting
  - Overlap with other IPPs and purchased data assets?
- Scope of capital formation
  - What costs to include and how much use of output?
- Prices and depreciation
  - Are they different than software and databases?

Outline

- National Accounts Background on Data Assets
- Methodology
- Experimental Results
- Conclusions and Next Steps

Overview of the Methodology

- Own-account data for the U.S. business sector

- Sum of costs = labor + capital + intermediate inputs
  - Consistent with other U.S. own-account IPPs and DZTT
  - Data-related activities: collect, store, analyze, manage

- Methodology sequence
  - Estimate time use allocated to data-related activities by occupation (Blackburn 2021)
  - Estimate wages by occupation and industry
  - Apply a markup to estimate the full sum of costs
  - Adjust for multiple counting and capital formation

Summary of the Estimation

- Production cost: $C_{\omega,i,t} = \alpha \sum \tau_{\omega} W_{\omega,i,t} H_{\omega,i,t}$

- Time use: $\tau_{\omega} = \rho_{\omega} s^{*}_{\omega}$ (Blackburn 2021)
  - ρ_{ω}: fraction of jobs engaged in at least 1 data-related activity
    - 203 data-related skills identified in a skills data set
    - Skills data set from Burning Glass Technologies (BGT)
  - s^{*}_{ω}: share of time allocated to data-related activities
    - Step 1: Landmark occupations – occupations with highest rate of employees engaged in data-related activities (20 of these)
    - Step 2: Non-landmark occupations – cosine similarity to the closest landmark occupation (> 900 of these)
      - Model fitted using doc2vec implementation
      - Online job ads from BGT

Summary of the Estimation

- Production cost: $C_{\omega,i,t} = \alpha \sum \tau_{\omega} W_{\omega,i,t} H_{\omega,i,t}$

- Wages
  - $W_{\omega,i,t}$: average annual wage data from BLS OEWS
  - $H_{\omega,i,t}$: annual employment data from BLS OEWS

- Markup (α)
  - Supplements, capital cost, and intermediate consumption
  - Data from BEA supply-use tables

- Adjustments applied by industry or by occupation

Landmark Occupations

Description	Time-Use Factor
Data Entry Keyers	0.94
Computer and Information Research Scientists	0.77
Database Administrators	0.75
Database Architects	0.72
Bioinformatics Scientists	0.68
Social Science Research Assistants	0.67
Statisticians	0.66
Data Warehousing Specialists	0.63
Biostatisticians	0.63
Business Intelligence Analysts	0.61
Wellhead Pumpers	0.60
Survey Researchers	0.59
Bioinformatics Technicians	0.58
Statistical Assistants	0.54
Hearing Aid Specialists	0.54
Clinical Data Managers	0.54
Statement Clerks	0.50
Methane/Landfill Gas Generation System Tech.	0.47
Geographic Information Systems Technicians	0.44
Intelligence Analysts	0.43

Markup (α)

- Weighted composite ratio from BEA supply-use tables
  - Data processing and hosting (NAICS 518)
  - Computer systems design (NAICS 5415)
  - Administrative services (NAICS 561)

	Ratio	Share
Compensation	1.17	48%
Intermediate consumption	0.78	32%
Consumption of fixed capital	0.18	8%
Net operating surplus	0.29	12%
Markup	2.42	100%

Adjustments

- Multiple counting
  - R&D: assume 50 percent for R&D intensive industries
    - Chemical manufacturing (NAICS 325)
    - Computer and electronics manufacturing (NAICS 334)
    - Transportation equipment manufacturing (NAICS 336)
    - Software publishing (NAICS 511)
    - Professional, scientific, technical (PST) services (NAICS 541)
  - Software: remove own-account software occupations
    - Computer programmers
    - Computer systems analysts
    - Software developers and software quality assurance analysts
  - Purchased data: assume 50 percent for PST services

Adjustments

- Capital formation (use of output)
  - Assume 50 percent for all industries

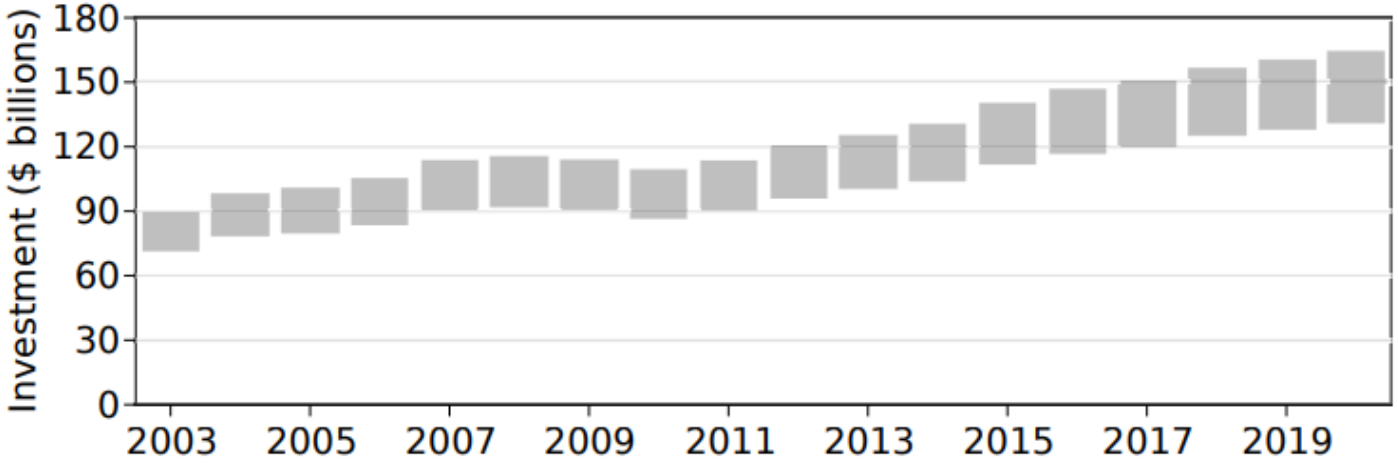
- 50 percent is a placeholder

NAICS	Markup	Capital formation	R&D	Purchased data	Eff. factor
325	2.42	0.50	0.50	N/A	0.605
334	2.42	0.50	0.50	N/A	0.605
336	2.42	0.50	0.50	N/A	0.605
511	2.42	0.50	0.50	N/A	0.605
541	2.42	0.50	0.50	0.50	0.3025
All other	2.42	0.50	N/A	N/A	1.21

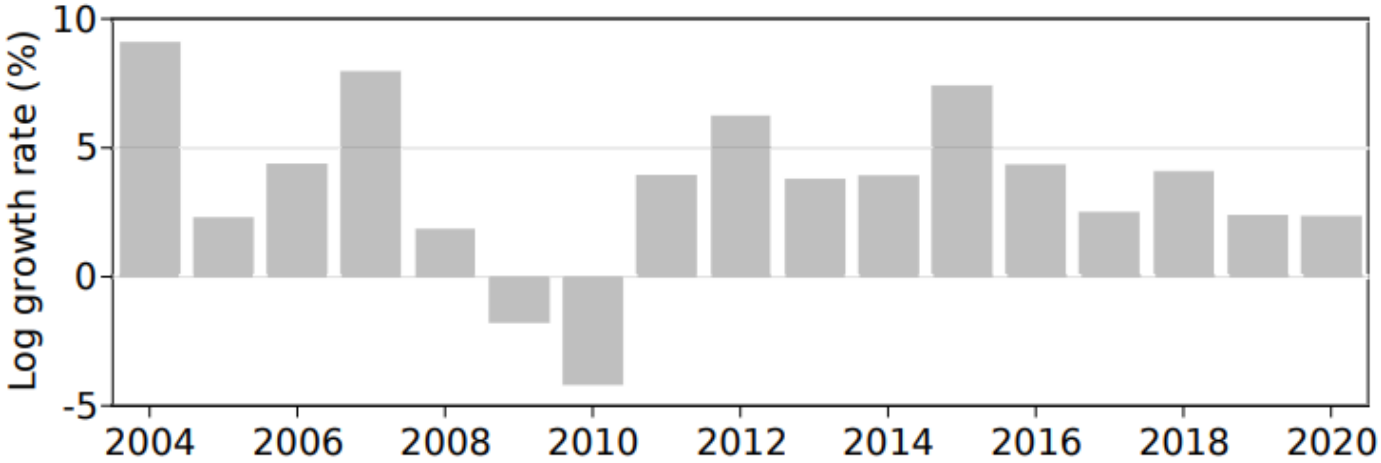
Outline

- National Accounts Background on Data Assets
- Summary of Methodology
- Experimental Results
- Conclusions and Next Steps

Current-Dollar Investment

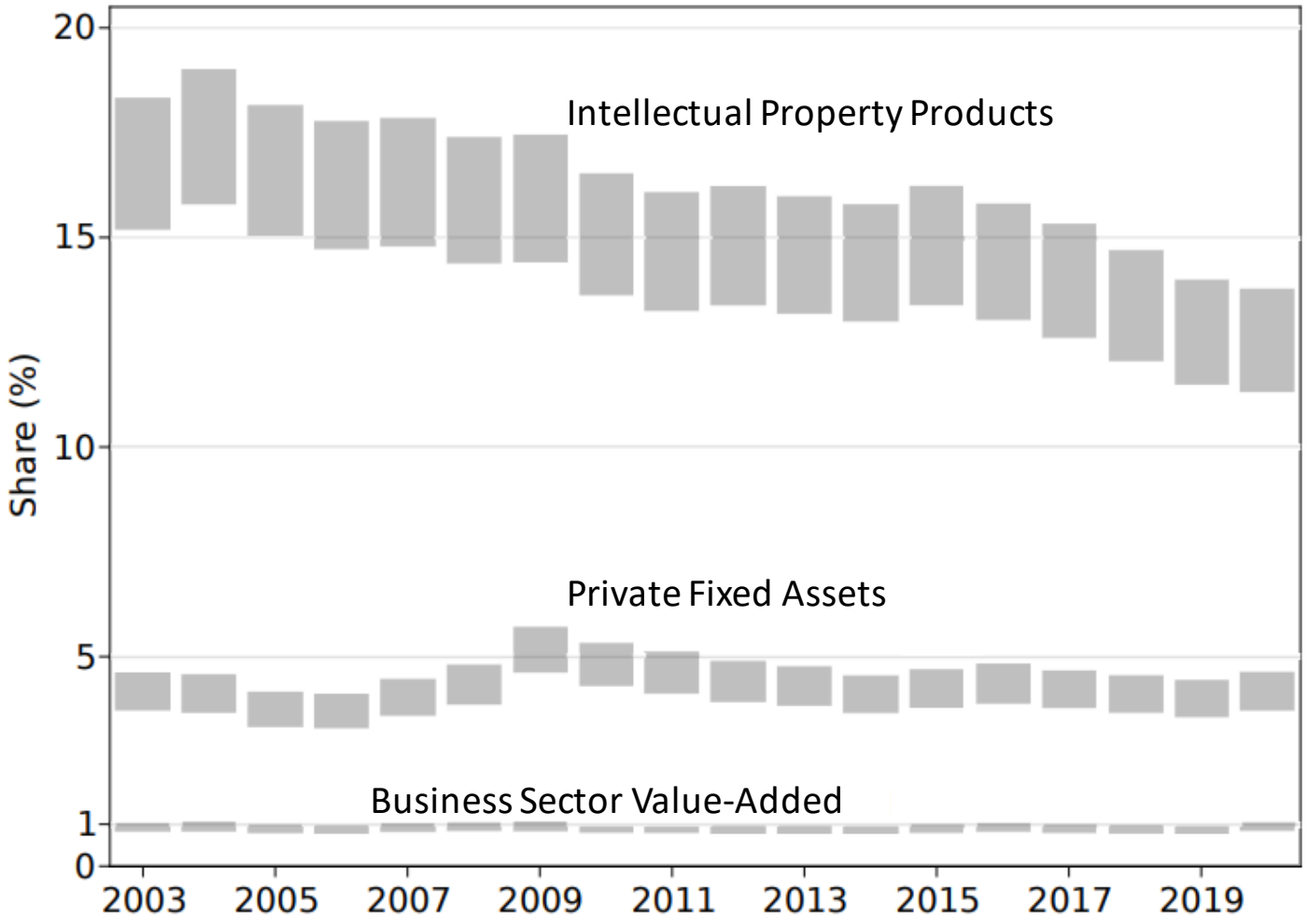


2003 = $72 B
2020 = $131 B



Average = 3.6%

Investment Shares of NIPA Aggregates

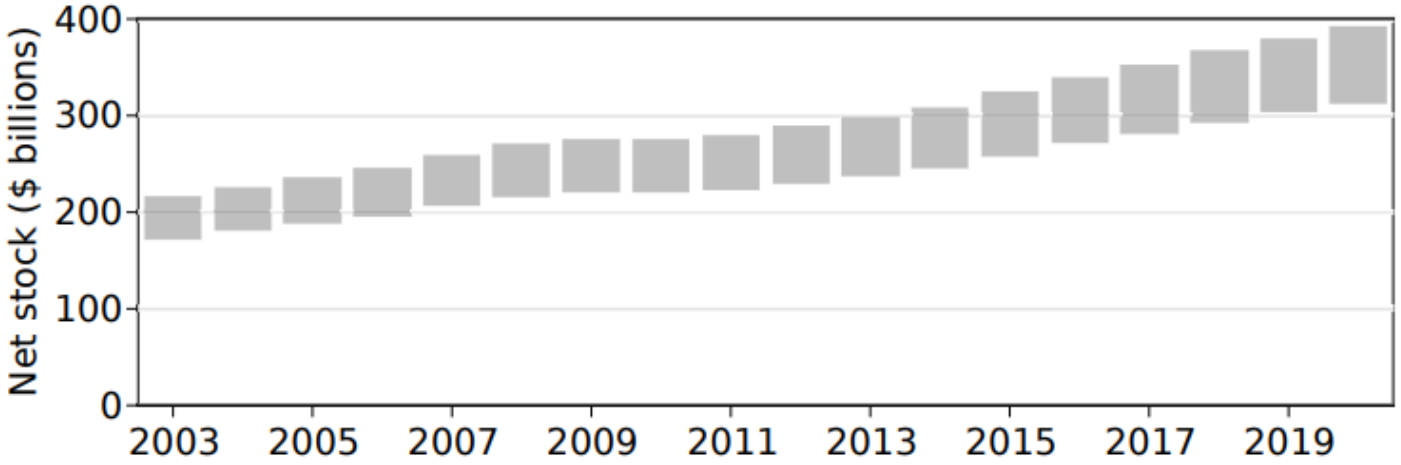


Average = 13.6%

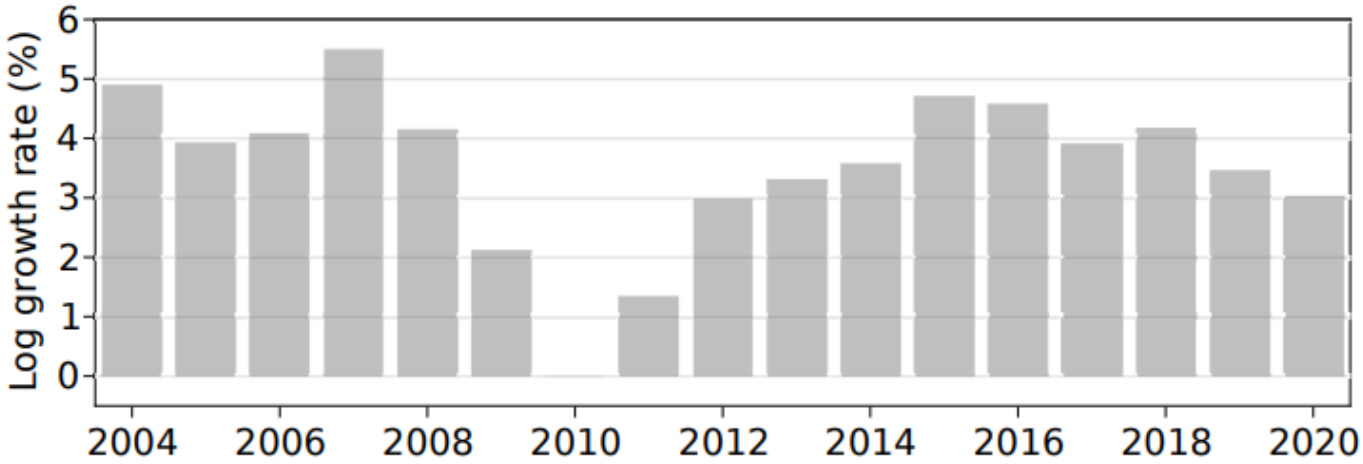
Average = 3.8%

Average = 0.8%

Historical-Cost Net Stocks

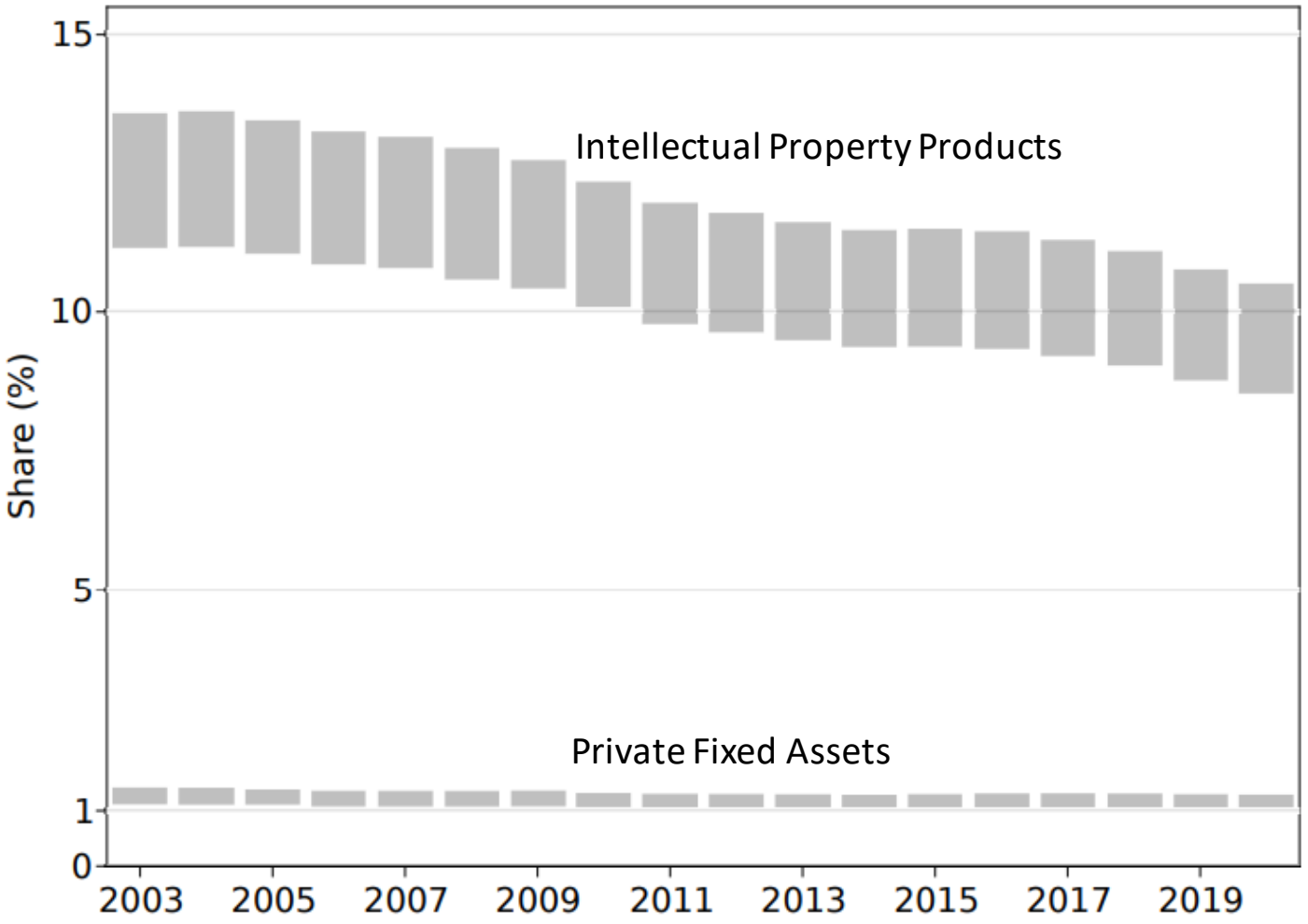


2003 = $174 B
2020 = $313 B



Average = 3.4%

Net Stock Shares of FAA Aggregates



Average = 10.0%

Average = 1.1%

Average Growth in Real Investment 2004-2020

	With data	W/o data	Δ
Data	3.76		
Business Sector Value-added	1.74	1.73	0.01
IPPs	4.83	5.06	−0.23
Software	6.42	7.58	−1.16

Current-Dollar Investment by NAICS 2003-2020

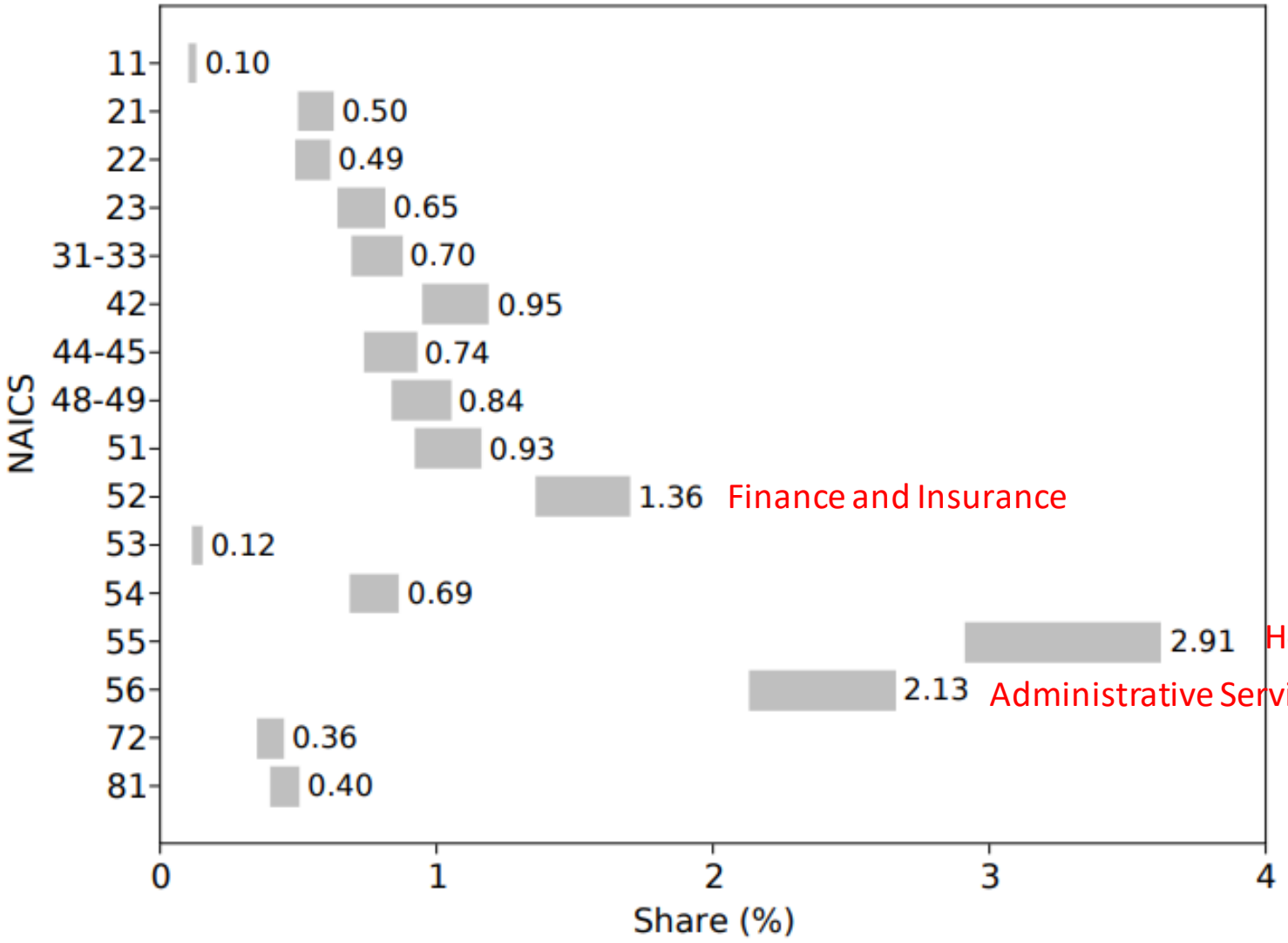
NAICS	Description	($B)
52	Finance and Insurance	301
31-33	Manufacturing	246
56	Admin. & Support and Waste Management & Remediation Services	192
42	Wholesale Trade	169
55	Management of Companies and Enterprises	162
54	Professional, Scientific, and Technical Services	149
51	Information	139
44-45	Retail Trade	127
23	Construction	80
48-49	Transportation and Warehousing	74
53	Real Estate and Rental and Leasing	46
72	Accommodation and Food Services	31
81	Other Services (except Public Administration)	27
21	Mining, Quarrying, and Oil and Gas Extraction	26
22	Utilities	25
11	Agriculture, Forestry, Fishing and Hunting	4
	Total	1,798

Current-Dollar Investment Growth by NAICS 2004-2020

NAICS	Description	(%)
55	Management of Companies and Enterprises	6.3
54	Professional, Scientific, and Technical Services	5.9
21	Mining, Quarrying, and Oil and Gas Extraction	4.6
53	Real Estate and Rental and Leasing	4.2
23	Construction	4.1
48-49	Transportation and Warehousing	4.0
11	Agriculture, Forestry, Fishing and Hunting	3.6
52	Finance and Insurance	3.6
56	Administrative & Support and Waste Management & Remediation Services	3.4
51	Information	3.3
81	Other Services (except Public Administration)	3.1
42	Wholesale Trade	3.0
22	Utilities	2.9
72	Accommodation and Food Services	2.7
31-33	Manufacturing	2.0
44-45	Retail Trade	1.9

3.6%

Current-Dollar Investment as a Share of Value-Added by NAICS 2003-2020



Share of Current-Dollar Investment by Occupation 2003-2020

Description	Share(%)
Office Clerks	6.3
General and Operations Managers	4.7
Data Entry Keyers	4.6
Database Administrators and Architects	4.5
Customer Service Representatives	3.9
Management Analysts	3.6
Bookkeeping, Accounting, and Auditing Clerks	3.5
Computer and Information Systems Managers	3.5
Secretaries and Administrative Assistants	3.0
Computer Occupations, All Other	2.6
Billing and Posting Clerks	2.4
Financial and Investment Analysts	2.2
Market Research Analysts and Marketing Specialists	2.2
Network and Computer Systems Administrators	2.0
First-Line Supervisors of Office and Administrative Workers	1.9
Financial Managers	1.7
Accountants and Auditors	1.7
Total	54.3

IPPs in the U.S. National Accounts, 2020

Total investment.....*$1,209.5*

Software, databases, and data.....584.4

- Software and databases.....453.4
  - Prepackaged.....$212.5
  - Custom.....165.5
  - Own-account.....75.4
- Data.....131.0
  - Purchased.....$???.?
  - Own-account.....131.0

Research and development.....537.7

Entertainment, literary, and artistic originals.....87.4

Source: NIPA Table 5.6.5. Billions USD.

NPISH Investment in Data 2003-2020

NAICS	Description	($B)
61	Educational Services	150
62	Health Care and Social Assistance	318
71	Arts, Entertainment, and Recreation	21
813	Religious, Grantmaking, and Similar Organizations	45
	Total	534

Outline

- National Accounts Background on Data Assets
- Summary of Methodology
- Experimental Results
- Conclusions and Next Steps

Conclusions

- Current-dollar investment in own-account data assets for the U.S. business sector grew from $72 billion in 2003 to $131 billion in 2020, which yields an average annual growth of 3.6%.
  - Growth in data investment > business sector value-added.
  - Growth in data investment << investment in other IPPs.
  - Growth by NAICS sector consistent with anecdotal evidence.

- Investment by occupation suggests that overlap with R&D and software investment seems to be mitigated.
  - Overlap between data landmark occupations and software-related occupations suggests data and software should be jointly estimated.

- Method for identifying occupations and estimating the time-use that occupations allocate to data-related activities is feasible.

Next Steps

- Work-in-progress
  - Develop an input-cost price index
  - Robustness for time-use and representative industries
  - Apply the method for identifying occupations and estimating time-use factors to potentially expand own-account software

- Future work
  - Expand scope to NPISH and government sectors
  - Estimate or identify more precise adjustments for purchased data assets, capital formation, and R&D and software overlap
  - Estimate or identify a data-specific service life/depreciation rate (≈ IDC Global StorageSphere)

Questions for Discussion

- Would a separate series on data as an asset be useful in the NIPAs?
- Have we missed any measurement issues?
- How can we vet and improve the adjustments?
- Do the recommendations proposed for data as an asset in *SNA 2025* seem reasonable?