

Toward Economics Statistics On: Own-Account Data and Databases Assets for 1997–2024*

Authors	José Bayoán Santiago Calderón, U.S. Bureau of Economic Analysis
Contact	Jose.Santiago-Calderon@bea.gov
Date	March 2026
Abstract	This article describes the proposed methodology and experimental estimates for capitalizing own-account data as part of the updates to the System of National Accounts from the 2008 standard to the 2025 revision. The methodology explains how the U.S. Bureau of Economic Analysis estimates annual industry-level nominal values, prices, and volume for gross fixed capital formation, consumption of fixed capital, and net stock for a new asset class: own-account data and databases. Additionally, it illustrates the impact of the change across the National Economic Accounts. The work spans the period from 1997 to 2024. The economy grew at an average rate of seven basis points faster from 2002 to 2024 when own-account data and databases were capitalized.
Keywords	Data Assets, System of National Accounts, IPPs, Digital Economy
JEL Code	E01, E22, O34, O51

* We are grateful for the support and feedback from the IMF-Eurostat Task Team and participant national statistical offices. The work benefited from the expertise and support of the BEA directorate of National Accounts leadership and the Investment section.

1. Introduction

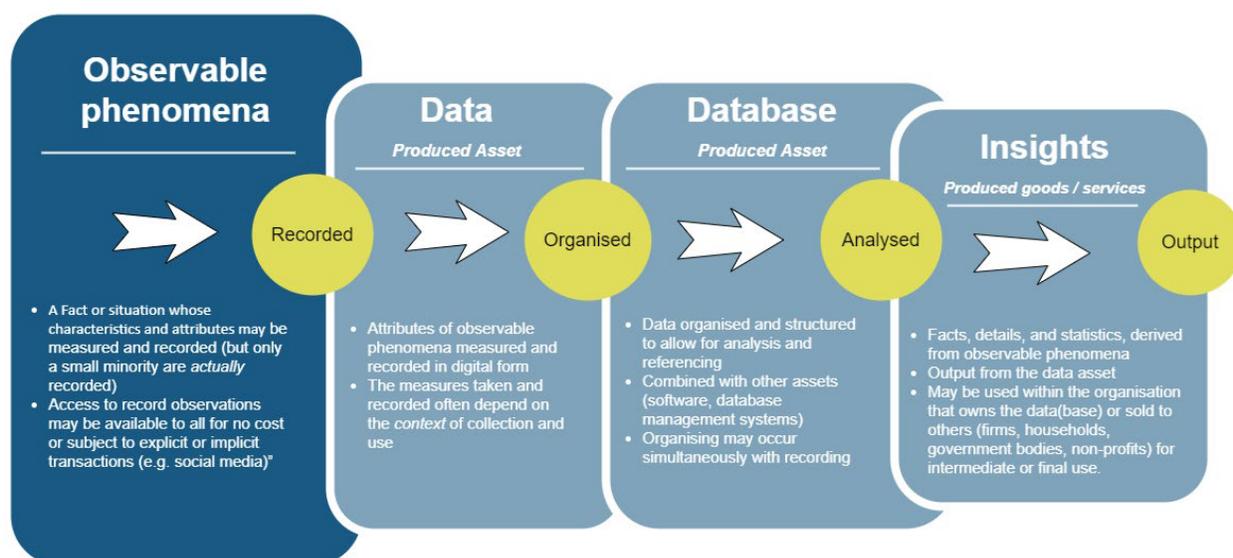
This article describes the proposed methodology used by the U.S. Bureau of Economic Analysis (BEA) for measuring annual industry-level nominal values, prices, and volume for gross fixed capital formation (GFCF), consumption of fixed capital (CFC), and net stock estimates for a new asset class: own-account data and databases. The National Income and Product Accounts implement the *System of National Accounts (SNA)* 2008 version (United Nations 2009), which does not consider own-account data as an asset. The *SNA 2025* updated the treatment of own-account data to capitalize it, that is, to treat spending in data as an investment rather than an expense (Intersecretariat Working Group on National Accounts 2025, §22.21–§22.32). Treating data as an asset acknowledges that organizations invest resources to generate and curate data that is expected to yield sustained future revenue streams and support ongoing production. Moreover, this change addresses an inconsistency in the treatment of data in capital formation depending upon whether a database is developed for own use or for sale or license (Rassier, Kornfeld, and Strassner 2019). The United Nations guidance for national statistical offices to implement the changes has a target window of 2029–2030. To position itself to implement the upcoming change to the national accounts framework, BEA is developing experimental economic statistics to preview the proposed methodology and collect feedback on it and the associated estimates.

BEA has used experimental economic statistics in the past when capitalizing other intellectual property products (IPPs), such as software in 1998 and research and development (R&D) in 2013 (Moylan and Okubo 2020; Parker and Grimm 2000). The methodology presented for these experimental statistics builds on a years-long research agenda at BEA. Santiago Calderón and Rassier (2025) introduced a general approach and initial estimates to address the updated treatment of data as an asset in the U.S. economy. Such efforts assisted the international measurement community and other national statistical offices in developing methodologies and experimental estimates (Coyle and Manley 2024; Schmidt, Pilgrim, and Mourougane 2023; Statistics Canada 2019). BEA received feedback from that work, which has been incorporated into the methods and estimates presented in this study. These changes include adjusting the production boundary to exclude data use (e.g., analysis), refining multiple-counting adjustments, and harmonizing definitions with those proposed in international standards (Advisory Expert Group on National Accounts 2023). Other improvements relative to prior work include the inclusion of government investment, extended time coverage, and treatment of self-employment, among others. Many of these lessons learned during the process and collaborations are discussed in the *Handbook on Measuring Data in the System of National Accounts* (Eurostat 2026).

1.1. Asset boundary

For the purposes of national accounts, data are digital records of information on observable phenomena, and databases are organized collections of data that enable resource-efficient use. In other words, as shown in figure 1, data are produced through the recording of observable phenomena stored in digital format that, when organized efficiently, produce databases. Data and database assets, by definition, must be repeatedly or continuously used in production for economic benefit (e.g., selling, licensing, self-use) for more than a year. Furthermore, the economic benefit must not serve another specific purpose or process, such as R&D, mineral exploration, or entertainment, as otherwise it would be captured under that corresponding account. It is worth noting that, in the national accounts, analog recordings are not considered data until they are “produced,” such as through digitization. The use of data, such as for analysis, is distinct from the production of data or databases, but it is critical to classify data and databases as assets. Lastly, data assets from data exhaust, resulting from the byproducts of core activities (e.g., transactions, payroll, financial statements), are limited to data that are intentionally stored or organized for use in other contexts. For example, payroll data are generated to pay employees but could also be repurposed to develop compensation packages to increase productivity or reduce costs. Sales data are generated to fulfill orders but can also be analyzed to inform logistics and pricing strategies.

Figure 1. Data-information Chain From a SNA Perspective



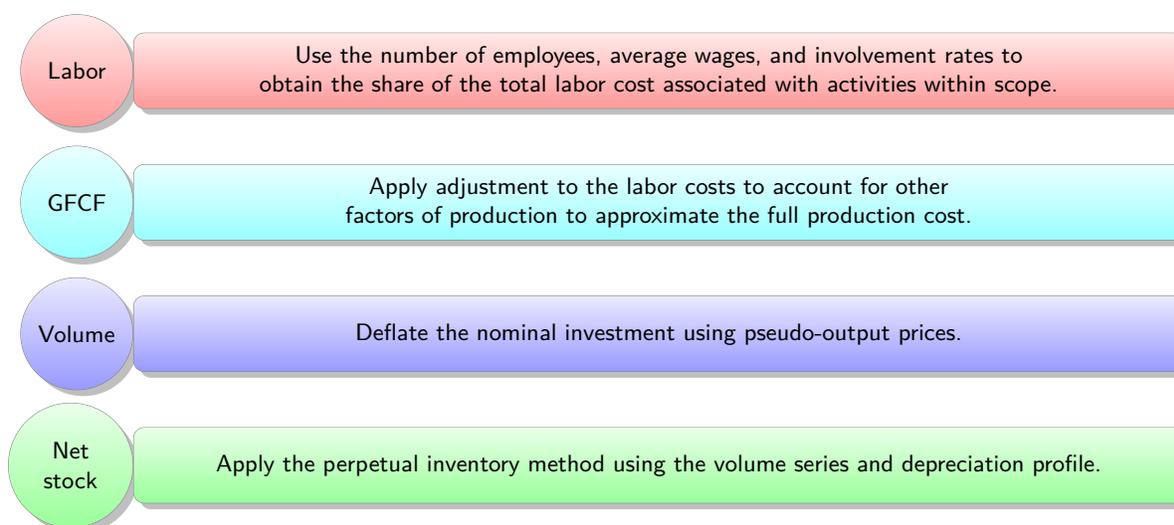
Source: Mitchell, Ker, and Leshar (2021)

1.2. Overview

The economic statistics for own-account data and databases will provide a comprehensive overview of the new asset class, encompassing annual industry-level estimates of nominal value, prices, and volume of gross fixed capital formation, consumption of fixed capital, and net stock. A subsequent release will expand the time coverage to include the historical time series.

Similar to own-account software, BEA uses the recommended sum-of-costs (SoC) approach to value own-account data and databases. The SoC approach calculates GFCF by first estimating the associated labor costs, which involves identifying the time effort allocation of labor towards GFCF and their corresponding remuneration. It then accounts for other costs using labor costs and a representative production function. Figure 2 depicts an overview of the strategy for each step towards fully accounting for the asset class in the National Economic Accounts framework.

Figure 2. Full Accounting for Asset Category Using the Sum of Costs Approach

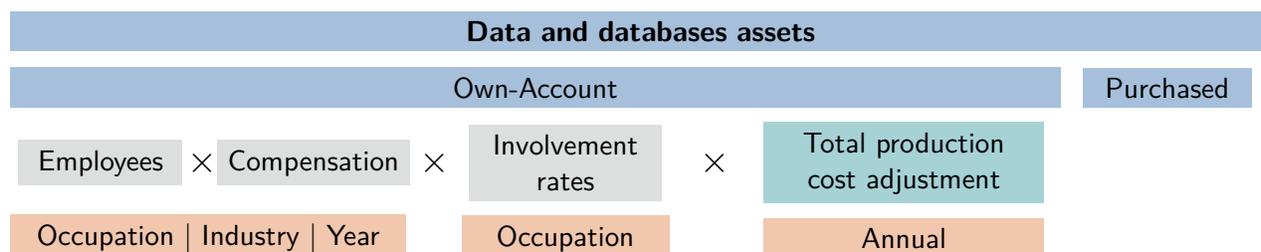


2. Gross fixed capital formation

Figure 3 illustrates the methodology for estimating own-account data and databases using the GFCF. It begins with obtaining annual industry-level estimates of employment and compensation. For each occupation, we estimate an involvement rate that, on average, quantifies the extent to which employees in that occupation contribute to the creation of data and database assets as part of their work activities. The product of the number of employees, their average wages, and the involvement rate, also known as the time-use factor or time-effort allocation, yields the share of wages and salaries contributing to GFCF. The share of wages and salaries is applied to total labor cost, which includes employee compensation and the labor share of proprietors' income, by industry, to obtain industry-level labor cost totals.

Labor costs are but one input in the production of the fixed asset. Other costs associated with producing the assets include capital costs and intermediate inputs. The total production cost adjustment, also known as the blow-up factor, accounts for these factors by applying a production function as a ratio. The resulting estimate represents the annual industry GFCF in data and database assets.

Figure 3. Estimating Gross Fixed Capital Formation



Consider equation 1 where for each occupation ω , industry i , and year t we calculate the associated payroll by multiplying the annual number of employees ($E_{\omega,i,t}$) by their average annual wage ($W_{\omega,i,t}$) and an occupation-specific involvement rate (τ_{ω}) that reflects the time-effort that the employees in that occupation allocate to data and database capital formation activities. The parameter α_t is an adjustment factor that reflects the cost of additional factors of production.

$$C_{i,t} = \alpha_t \sum_{\omega} \tau_{\omega} W_{\omega,i,t} E_{\omega,i,t} \quad (1)$$

3. Labor costs

This section explains how to obtain and use the necessary components to estimate labor costs for the GFCF of data and database assets, including compensation and involvement rates.

3.1. Compensation

The first step is to allocate payroll across the various occupations within each industry. BEA uses the U.S. Bureau of Labor Statistics (BLS) Occupational Employment and Wage Statistics (OEWS) to obtain the distribution of employee compensation from the national accounts by industry, disaggregated by occupation groups (U.S. Bureau of Labor Statistics n.d.). The OEWS program annually produces industry-specific and by-ownership employment and wage estimates for the United States for approximately 830 occupations (*ibid.*). The estimates include the number of employees and the average hourly wages and salaries by industry. However, the program is a cross-sectional data product, which presents various challenges when used as a time series. Relevant changes to the data include: (1) occupations (Hopson

2021), (2) industries (Fort and Klimek 2018), (3) sample design, (4) estimation methods, (5) data collection, (6) scope, and (7) reference period. With those considerations, BEA's approach is to (1) identify time-consistent occupation groups to aggregate the estimates to that level and (2) assume an equal number of hours worked by occupations within an industry. These assumptions allow the use of the cross-sectional data as a time series. However, some limitations persist, including limited coverage of the farming sector and partial coverage of forestry, fishing, and related activities.

The OEWS data used include annual estimates of the number of employees and their average hourly wages by industry, using the North American Industry Classification System (NAICS), and establishment ownership type (i.e., private, state and local, federal). The survey does not provide additional information on employees' hours worked (e.g., full-time vs. part-time or seasonal). We aggregate payroll estimates by multiplying the number of employees by the average hourly wage and summing across three-digit and four-digit NAICS categories to match the summary industry detail accounts. The OEWS uses the NAICS classifications for estimates beginning in 2002 and the Standard Industrial Classification (SIC) taxonomy for 1997–2001. The SIC estimates are mapped to NAICS and the summary industry accounts using BLS crosswalks that use the share of employment allocated from SIC to NAICS at the time of the transition. For the years post-2007, when OEWS reports ownership for education and healthcare (NAICS 61–62), the employees/payroll attributed to state and local or federal government are reclassified to the corresponding government account, which includes public administration (NAICS 92) and the postal service (NAICS 491) for the federal government.

BEA also computes a normalized cost per unit of labor (e.g., employee, full-time equivalent, hours worked) to distinguish between quantity and cost per unit, which will be used when developing prices for the asset category. Table 1 presents an example of ideal data where, for each detailed occupation in each industry, for each year, it provides an estimate of the employee compensation per unit of labor and how many units of labor were employed. In practice, it would be rare, if not impossible, to obtain such detailed data.

Table 1. Example of Desiderata for Labor Costs

SOC	SOC title	NAICS	Year	Hourly employee compensation	Hours
19-3011	Economist	521	2022	\$103	200
19-3011	Economist	521	2023	\$105	180

BEA estimates a time series of the annual share of compensation for employees and the number of units of work performed by each occupation, based on a set of assumptions and procedures. It requires generating a directed graph from the crosswalks between the occupational taxonomies in the OEWS tables and identifying the graph subcomponents (i.e., all occupation codes that are interconnected in a given vintage). New series are generated by aggregating common occupation groups, computing the

aggregate payroll (i.e., the number of employees multiplied by their average wages), and dividing it by the new aggregate number of employees. Using the new values, the estimated industry shares of employment and payroll are applied to employee compensation in the National Economic Accounts to derive annual compensation estimates by industry and occupation. The process of going from wages and salaries to labor cost is shown by equation 2 where W represents the wages for occupation ω in industry i in year t , E the number of employees, τ the involvement rate, and Λ the labor cost.

$$L_{\omega it} = \Lambda_{it} \frac{\tau_{\omega} W_{\omega it} E_{\omega it}}{\sum_{\omega it} W_{\omega it} E_{\omega it}} \quad (2)$$

The labor cost is obtained from the BEA-BLS Integrated Industry-level Production Account (ILPA) (U.S. Bureau of Economic Analysis 2025). The labor cost includes employee compensation, the share of the proprietor's income attributed to labor (i.e., self-employment income), and adjustments for taxes and subsidies that affect the true cost of labor. In essence, this allocates the total measured labor cost in each industry among occupations according to their wages. It then calculates the fraction of compensation for each occupation that is considered to contribute to data capital formation and uses labor costs as a component of a production function to determine total production cost. When the ILPA publishes aggregate-level labor compensation estimates that do not match the industry-level summary, we assume that the distribution of labor compensation follows that of employees' compensation in the gross GDP by Industry/KLEMS published estimates.¹

To obtain the changes in price per unit (i.e., average compensation for a unit of work), we explored a few alternatives, such as using the estimated number of hours in each industry from BEA's National Economic Accounts and BLS' productivity office (U.S. Bureau of Economic Analysis 2023b,c; U.S. Bureau of Labor Statistics 2024a). The proposed methodology assumes that employees within the industry work similar hours. However, to develop prices, we need to aggregate labor costs across sectors by unit of work. For this purpose, we estimate the share of total time under the assumption that workers across occupations within an industry work similar hours and apply this share to the hours worked in each industry, allowing for differences across industries (U.S. Bureau of Labor Statistics 2024a).

The unit of cost of labor is given by equation 3 where L is given from equation 2, E is the number of employees from OEWS, and H is the number of hours worked in industry i in year t . This assumption allows for differences, such as industries with higher rates of part-time employees, employees working more or fewer hours in an industry than in other industries, and varying ratios of hours between part-time and full-time employees across industries.

¹The BEA Gross Domestic Product by Industry and the KLEMS releases provide estimates of the factors of production, capital, labor, and intermediate inputs for each industry.

$$P_{Lt} = \sum_{it} \frac{L_{it}}{H_{it} \sum_{\omega_i} \tau_{\omega} \frac{E_{\omega_{it}}}{E_{it}}} \quad (3)$$

3.2. Involvement rates

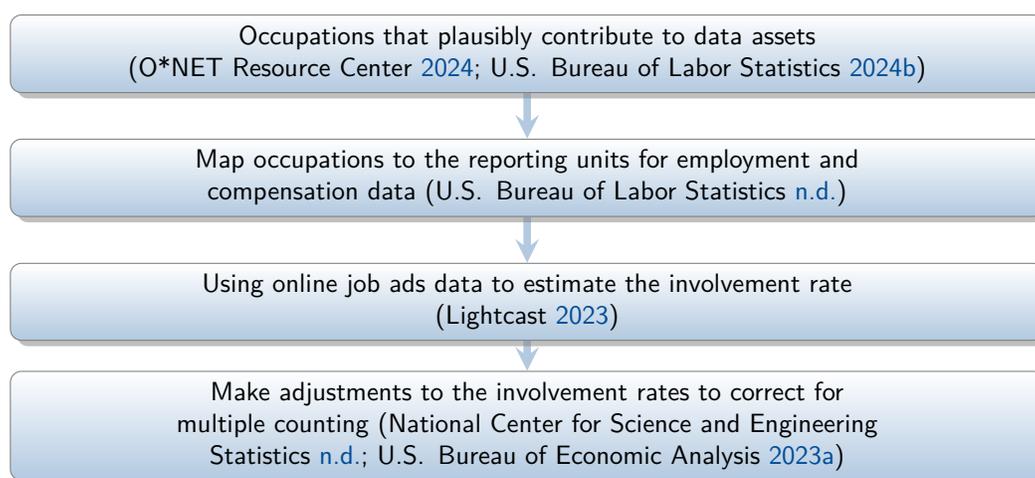
BEA employs a mixed approach for estimating the involvement rates associated with the SoC approach depicted in figure 4. For data and database assets, BEA first uses content models to assess occupations likely to contribute to their production through core work activities. The O*NET content model employs a theoretical framework to identify and organize information about work, utilizing surveys and expert knowledge. For example, it estimates the educational level of workers in each occupation, the types of skills they use, the contexts in which they use them, the extent to which those skills are used, and the degree to which they are relevant to performing various tasks and work activities. This approach considers the work and detailed work activities for each occupation, as well as the skills (frequency, relevance, depth) and work contexts, to determine the significance of that occupation to the activities of interest (O*NET Resource Center 2024). The work context facilitates addressing various measurement issues, such as capturing work that produces data but not data assets, because the data are short-lived or occur in the context of R&D or mineral exploration, which is not the target. These are further evaluated based on the descriptions of their roles and duties in the BLS Occupational Outlook Handbook (U.S. Bureau of Labor Statistics 2024b). The main job description and role also help determine whether certain data-generating work activities are intentional or the result of data exhaust, thereby enabling cleaner estimates of involvement rates under the operationalized definition. In summary, BEA uses expert judgment to preselect a list of occupations likely to generate data assets.

Context also helps exclude data that are unlikely to be own-account data assets, such as occupations that primarily involve the formation of data assets for sale, licensing, or that are not used internally. For example, clinical laboratory technologists and technicians likely produce data, but it is unlikely to constitute own-account data assets. However, the original data assets that support the sale or licensing of copies would fall within the scope of own-account data. While the context could be tailored to specific occupation-industry combinations, we conduct the analysis at the occupation level, accounting for industry distribution.

The second approach, beyond content models, is the use of machine learning to extract insights from large datasets. Once a consistent taxonomy of occupations that matched payroll estimates had been identified, BEA used the Lightcast online job-advertisement dataset to generate a representative sample of job ads (Lightcast 2024). The sampling design examines occupations with sufficient observations to compute estimates of the share of jobs that are engaged in specific work activities. The sampling design uses the provided variable in the dataset “cleaned job title” which is recoded using the Occupational

Information Network (O*NET Resource Center 2024) autoencoder to classify it into an O*NET code, which is then mapped to the occupation group that matches the payroll data. The sampling is designed to mirror the occupational distribution across industries in the multi-year average of payroll data (the average of the 2015–2020 distribution of jobs across industries). Using this sample, BEA estimates the share of job ads that list one of the identified skills, which proxies a work activity that contributes to data and databases GFCF (Lightcast 2023). The relevance of skills to data and database production was determined manually by examining the skills and their roles in activities likely to produce data and database assets (e.g., digitization, database management). The skills are very similar to those in prior work (Santiago Calderón and Rassier 2025), but are refined to exclude the use and analysis of data under the updated guidance. Only occupations that have been identified as probably contributing to GFCF are included, and every other involvement rate estimate is set to zero.

Figure 4. Mixed Approach for Involvement Rates



The estimated involvement rates are then rounded to the nearest 5%, with values below 10% set to 0, and estimates above 90% capped at 90%. There are two explicit adjustments for multiple counting: R&D and software. The R&D adjustment is based on the estimated rate of the variable “primarily engaged in R&D” by occupation group based on the National Survey of College Graduates (National Center for Science and Engineering Statistics n.d.). For example, if for occupation X the estimated share of employees who mainly reported performing R&D is 20% and the estimated involvement rate for GFCF of data and databases for that occupation was 40%, then the R&D adjusted involvement rate for GFCF in data and database assets would be $(40\%)(1 - 20\%) = 32\%$. For software, the adjustment uses the latest involvement rates for own-account software activities applied in the same manner. The final involvement rates are then applied to the occupation payroll, divided by the total payroll, to obtain the final shares of wages and salaries in each industry attributed to GFCF.

4. Full production cost adjustment factor

Labor compensation is only a portion of the cost of producing data. The total production cost adjustment factor seeks to align labor cost estimates with the total production cost measure (i.e., gross output).

4.1. Representative industries

A primary consideration is identifying a representative production function for the activity of interest. For data and databases, the selected representative industries are data processing, hosting, and related activities (ILPA 514) and computer programming, consultancy, and related activities (NAICS 5415).² The basic approach is to use the cost structure in these industries to scale labor costs for the production of data by industry to total costs.

4.2. Labor, capital, and intermediate consumption

BEA develops the full production cost adjustment factor by considering the representative industries as a whole. It uses equation 4 to compute the ratio of the sum of gross output (at basic prices) for industries 514 and 5415 to the sum of compensation for non-college workers and compensation for college workers across both industries. The underlying assumption is that total costs equal gross output; that is, labor, capital, and intermediate inputs equal gross output.

$$\text{Full Production Cost Adjustment Factor}_t = \frac{\sum_{j \in \{514, 5415\}} \text{Gross output}_{jt}}{\sum_{j \in \{514, 5415\}} \text{Compensation of Labor}_{jt}} \quad (4)$$

Let the total production cost adjustment factor be denoted by a_t . Then, to compute the value applied to employee compensation at time t , t is given by equation 5, where the factor applied is the 10-year-lagged simple moving average. This same factor applies to all industries that employ workers who produce data.

$$\alpha_t = \frac{\sum_{t \in \{t-9, t-8, \dots, t\}} a_t}{10} \quad (5)$$

²The ILPA publishes account 514, which includes NAICS 518/519 together.

5. Prices

A pseudo-output price index is developed for the data and databases asset class. The price index uses three main components: labor, intermediate inputs, and a productivity adjustment. The labor cost component is estimated based on the labor compensation discussed above, divided by the effective number of hours, which is calculated by multiplying the share of effective time-effort across industries, weighted by the hours worked by workers in each industry (see equation 3). The movement in the cost of intermediate inputs is based on published industry-level prices for each component (energy, materials, and purchased services) for representative industries in the BEA industry accounts. The change in the cost per unit is weighted by the respective industries' share weights within the composite representative industries. This follows the standard price aggregation formula, which is a weighted sum of price changes, with weights proportional to the expense shares. The productivity adjustment is applied by subtracting the 5-year moving average of the change in total factor productivity for the two industries (weighted by gross output) from the labor and intermediate input series. The complete calculation for the log growth change in prices Δp_t is given in equation 6

$$\Delta p_t = \sum_{jct} \alpha_{ct} \Delta_{ct} - \frac{\sum_{t \in \{t-4, t-3, \dots, t\}} \sum_j \frac{GO_j}{\sum_j GO_j} [\ln(\rho_{jt}) - \ln(\rho_{j,t-1})]}{5} \quad (6)$$

where c represents a cost component: labor, energy inputs, material inputs, and purchased-services inputs for industry $j \in \{514, 5415\}$ at year t . Continuing with notation, α refers to the share weight of the component out of the total costs that year, as shown in equation 7.

$$\alpha_{cjt} = \frac{c_{jt}}{c_t} \quad (7)$$

Furthermore, ρ denotes the total factor productivity level for industry j at time t . Lastly, equation 8 defines Δ as the log growth change in prices for a component, which are taken from the BEA KLEMS tables for the intermediate inputs and the labor cost per unit from equation 3.

$$\Delta_{ct} = \ln(p_{ct}) - \ln(p_{c,t-1}) \quad (8)$$

6. Net Stock

The current-cost net stock series is estimated using the perpetual inventory method. The depreciation profile for the new asset class was set to a geometric depreciation rate of 1/3, based on an average service life of five years and a declining rate of 1.65, which is also used for software. The volume series for the asset class has been estimated using the initial investment divided by the depreciation rate as the seed in 1997, with ongoing efforts to estimate the historical series. This capital stock benchmarking is subject to change once additional estimates for the historical series are available.

7. Results

This section presents results from applying the methodology described in the previous sections from 1997 to the current year to obtain estimates of GFCF, prices, volume, CFC, and net stock.

7.1. Involvement rates

The involvement rates resulting from applying the methodology in section 3.2 are shown in tables 2 and 3. Note that some occupation codes are at the most detailed level, while others are at a broader level. This is an artifact of the level of detail available for the employment and wages data, or of the aggregation used to report it, depending on whether the estimated involvement rates were the same across all detailed occupations within a broad group. The cases with multiple occupations in the same row represent occupations that are reported together due to breaks in the reporting taxonomy. The involvement rates range from 10% to 90% by construction. However, involvement rates alone do not indicate the level of contribution to capital formation, as this also depends on the payroll share (i.e., the number of employees and average wages).

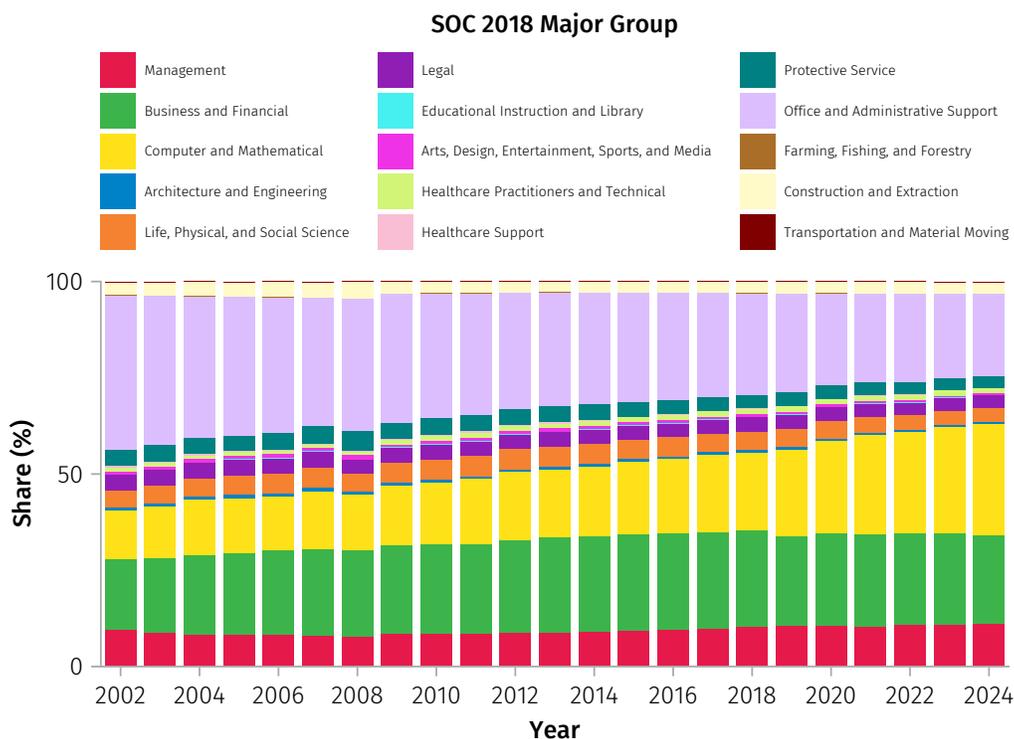
To assess which occupations contribute most to GFCF, we can examine the share of GFCF attributable to the Standard Occupational Classification (SOC) major occupational group. Figure 5 shows that the occupations with the most considerable contributions are: business and financial operations (SOC 13), computer and mathematical (SOC 15), and office and administrative support (SOC 43). A consistent second-order contribution is from management (SOC 11), in part due to the high compensation of these occupations. At the third order, there are life, physical, and social science (SOC 19), legal (SOC 23), and protective services (SOC 33). The legal and protective services occupations encompass activities in areas such as law enforcement, judicial, inspection, and safety. The relatively muted contributions of life, physical, and social science occupations are due to refinements that exclude data and database assets within R&D, which are already captured in the R&D account. Considering the most significant

contributions, it appears that there has been a shift in the types of data assets or in the asset-formation process from office and administrative support to computer and mathematical occupations over the last few decades.

The contribution of other major occupational groups is minor or negligible. For example, the architecture and engineering (SOC 17) group also faces a significant adjustment to avoid double-counting with R&D. Contributions from arts, design, entertainment, sports, and media (SOC 27) are affected by multiple-counting adjustments with the entertainment, literary, and artistic originals (ELAO). In the case of healthcare practitioners and technical (SOC 29), the approach was quite conservative in treating medical records and the service life of many diagnoses. For example, known allergies or the diagnosis of a chronic condition would have a long service life; many tests, such as imaging or lab work, tend to have a relatively short life or not be used repeatedly (e.g., need to be updated, serve as a basis to direct treatment for a particular instance). Clinical trials and other uses of medical data would be captured in the context of R&D. In terms of farming, fishing, and forestry (SOC 45), it seems various related activities are performed by occupations in other groups, and the activities performed by the major group tend to fail the definition of assets due to short service life. Construction and extraction (SOC 47) appears to make a visible contribution, in part due to the potential multi-year construction process and the long service life of structures, which facilitates meeting asset requirements. On the other hand, transportation and material-moving (SOC 53) occupations tend to generate records with a short service life, such as inventories that are unlikely to meet the requirement of more than a year.

It is desirable to only include occupations whose inclusion or exclusion affects the estimates of GFCF. For example, in figure 6, the cumulative contributions are shown by occupation count, revealing the annual minimum and maximum throughout the series. Between 2002 and 2024, including half of the occupations (approximately the top 35), accounts for approximately 95% of the GFCF in the series. The final set of occupations is denoted in the fourth column of tables 2 and 3.

Figure 5. Share of SOC Major Group Contribution



Note: The Standard Occupation Classification (SOC) of 2018 is used to summarize the occupational groupings.

Figure 6. Cumulative Contribution

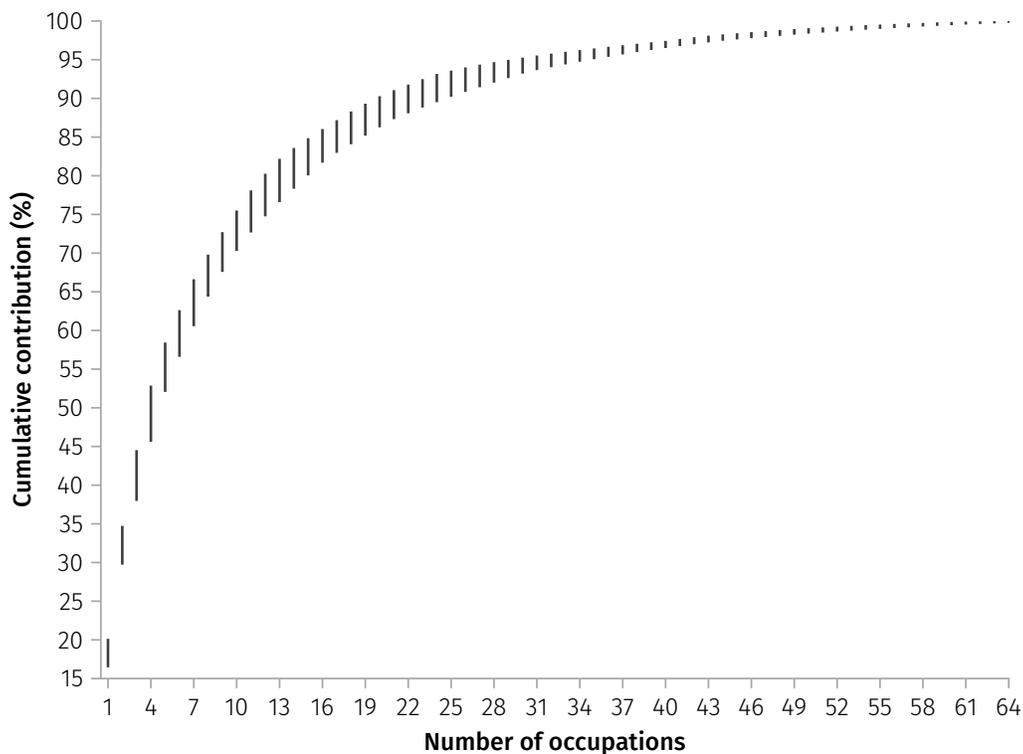


Table 2. Involvement Rates for Major Groups 11–19

SOC 2018	Title	%	Selected
11-3030	Financial managers	15	X
11-3050	Industrial production managers	10	X
13-1081	Logisticians	20	X
13-1160	Market research analysts and marketing specialists	10	X
13-2010	Accountants and auditors	30	X
13-2020	Property appraisers and assessors	15	X
13-2081	Tax examiners and collectors, and revenue agents	10	X
15-1242 15-1243	Database administrators; database architects	60	X
15-1252	Software developers	10	X
15-2010	Actuaries	10	
15-2030	Operations research analysts	20	X
15-2040	Statisticians	20	X
15-2050 15-2090	Data scientists; misc math sci occ	40	X
17-1021	Cartographers and photogrammetrists	25	
17-1022	Surveyors	10	
17-3030	Surveying and mapping technicians	15	X
19-1011	Animal scientists	10	
19-1013	Soil and plant scientists	10	
19-1022	Microbiologists	20	
19-1023	Zoologists and wildlife biologists	10	
19-1032	Foresters	10	
19-1041	Epidemiologists	45	
19-2032	Materials scientists	15	
19-2041	Env sci & hlth specs	20	X
19-2042	Geoscientists, except hydrologists and geographers	30	X
19-2043	Hydrologists	15	
19-3010	Economists	10	
19-3020	Survey researchers	45	
19-3032	Industrial-organizational psychologists	15	
19-3040	Sociologists	55	
19-3050	Urban and regional planners	10	
19-3092	Geographers	20	
19-3093	Historians	15	
19-4020	Biological technicians	35	X
19-4030	Chemical technicians	10	
19-4042	Env sci hlth & prot techs	15	
19-4050	Nuclear technicians	10	
19-4060	Social science research assistants	25	
19-4092	Forensic science technicians	10	

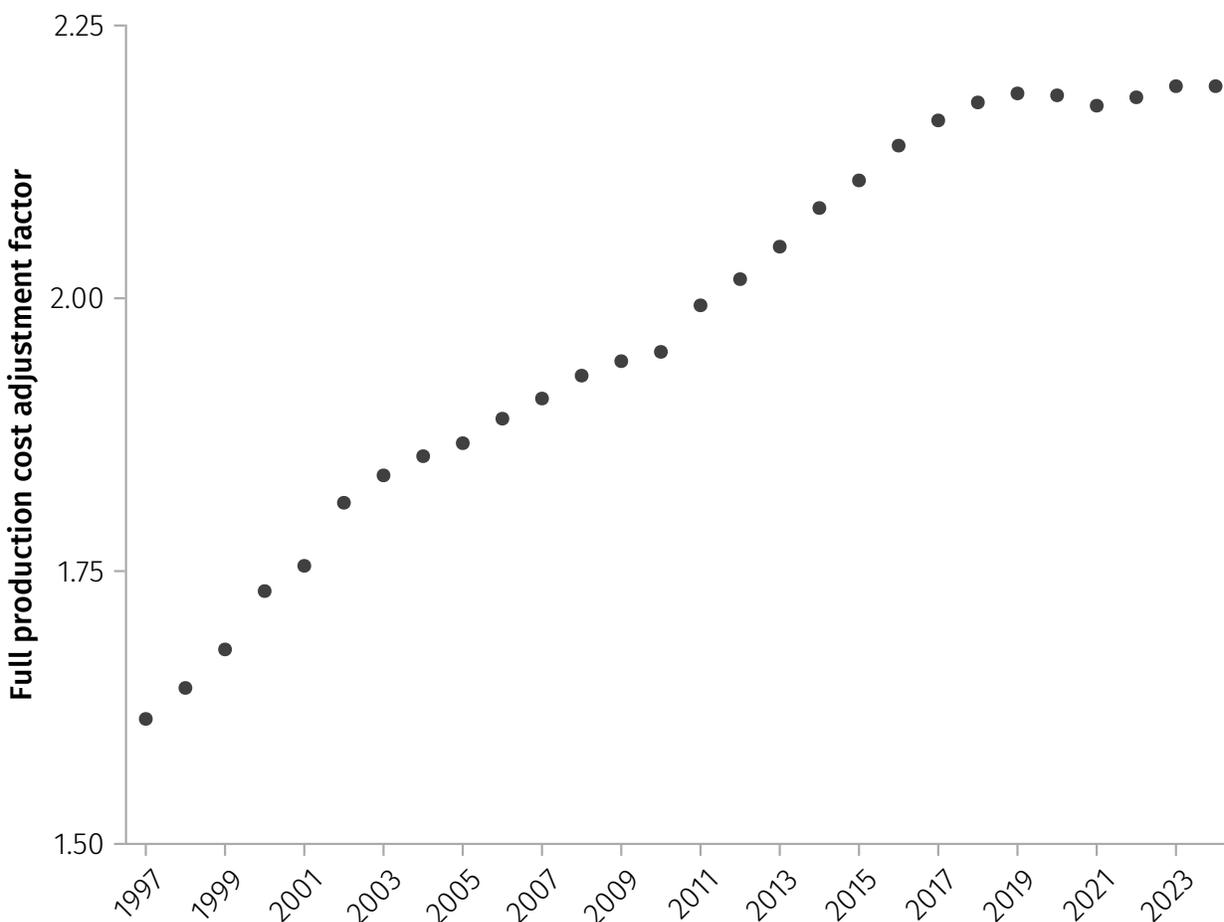
Table 3. Involvement Rates for Major SOC Groups 23–53

SOC 2018	Title	%	Selected
23-1012	Judicial law clerks	15	
23-1021	Administrative law judges, adjudicators, and hearing officers	35	X
23-1023	Judges, magistrate judges, and magistrates	10	X
23-2010	Paralegals and legal assistants	25	X
25-4010	Archivists, curators, and museum technicians	15	
27-3042	Technical writers	10	
27-3092	Court reporters and simultaneous captioners	60	X
29-2030	Diagnostic related technologists and technicians	10	X
29-2056	Veterinary technologists and technicians	10	
31-9094	Medical transcriptionists	20	X
33-2021	Fire inspectors and investigators	15	
33-2022	Forest fire inspectors and prevention specialists	45	
33-3020	Detectives and criminal investigators	40	X
33-3030	Fish and game wardens	15	
43-3030	Bookkeeping, accounting, and auditing clerks	25	X
43-3050	Payroll and timekeeping clerks	30	X
43-3060	Procurement clerks	15	X
43-4030	Court, municipal, and license clerks	30	X
43-4060	Eligibility interviewers, government programs	25	X
43-4070	File clerks	45	X
43-4110	Interviewers, except eligibility and loan	15	X
43-5040	Meter readers, utilities	15	
43-6012	Legal secretaries and administrative assistants	30	X
43-6013	Medical secretaries and administrative assistants	30	X
43-9021	Data entry keyers	90	X
43-9022	Word processors and typists	20	X
43-9040	Insurance claims and policy processing clerks	20	X
45-2010	Agricultural inspectors	10	
47-4010	Construction and building inspectors	60	X
53-2022	Airfield operations specialists	15	
53-6041	Traffic technicians	25	

7.2. Full production cost adjustment factor

The full production cost adjustment factor is estimated using the methodology described in section 4. Figure 7 shows the adjustment factor using the ILPA releases covering the available period starting from 1987 (U.S. Bureau of Economic Analysis 2022, 2025). The out-year (2024) is imputed using the last observation carried forward method.

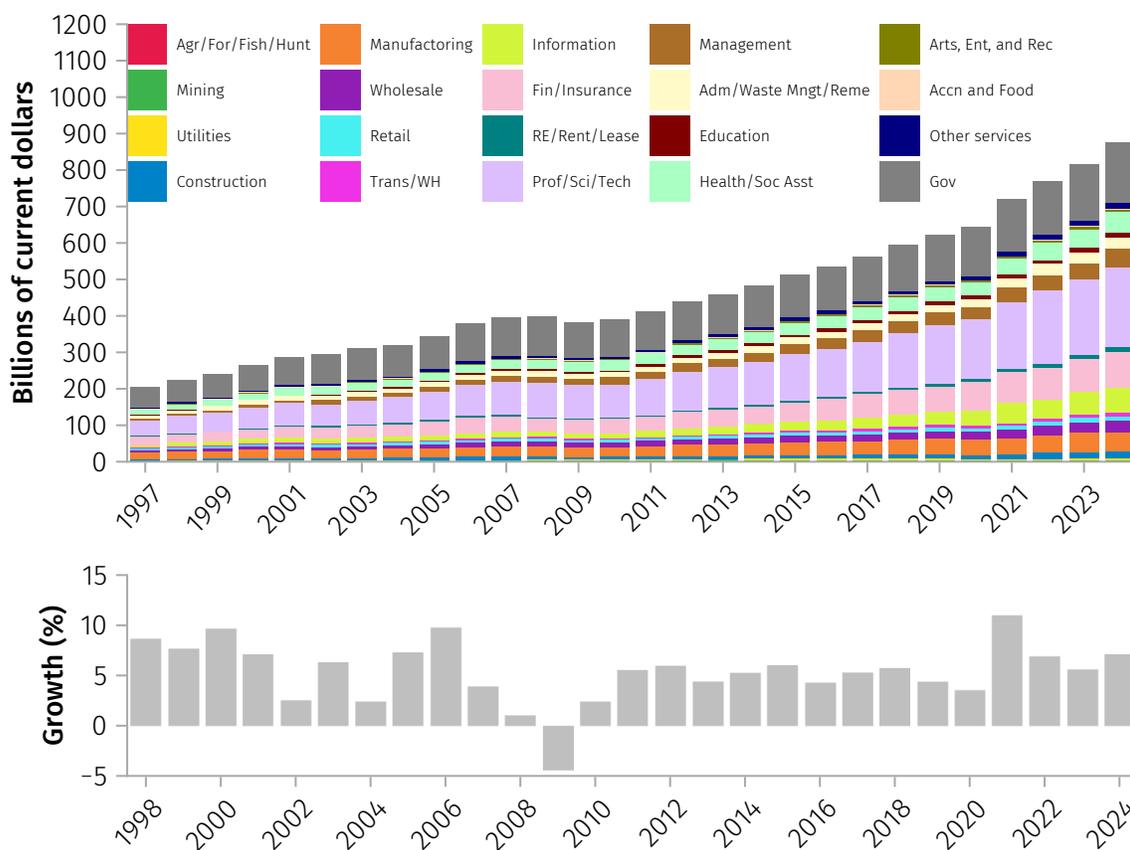
Figure 7. Full Production Cost Adjustment Factor



7.3. GFCF

The nominal investment series shows GFCF ranging from \$205 billion in 1997 to \$875 billion in 2024. Figure 8 shows the industries with the largest contributions, including professional, scientific, and technical services (PST) (ILPA 54); government (ILPA G); and finance and insurance (ILPA 52). At the second-order level, the series includes health care and social assistance (ILPA 62), manufacturing (ILPA 31G), information (ILPA 51), and management of companies and enterprises (ILPA 55). The nominal growth rates have been positive except during 2009, reflecting the Great Recession.

Figure 8. Gross Fixed Capital Formation



7.4. Prices

Estimating prices for the asset class yields the series shown in figure 9, alongside published GDP prices for reference. Prices increased from 1997 to 2001 and generally remained high until 2009. Prices increased more slowly than in the late 1990s and early 2000s, until the COVID-19 pandemic, when they declined. The most significant one-year decrease was -3% in 2009, and the increase was 8% in 2000, during the Dot-com bubble. The series bottomed at 85.97 in 1997 and peaked at 107.22 in 2024.

The price movement can be broken down into three components: labor, intermediate inputs, and productivity. Figure 10 shows that in the earlier years, prices experienced a faster growth in the cost of labor and negative productivity. Since 2002, productivity gains have largely contributed to the falling prices. The period 2021–2022 experienced higher-than-usual pressure on intermediate inputs, and 2024 experienced considerable pressure on labor costs. The productivity component for 2024 (outyear) is imputed from the preceding period's value.

Figure 9. Pseudo output prices

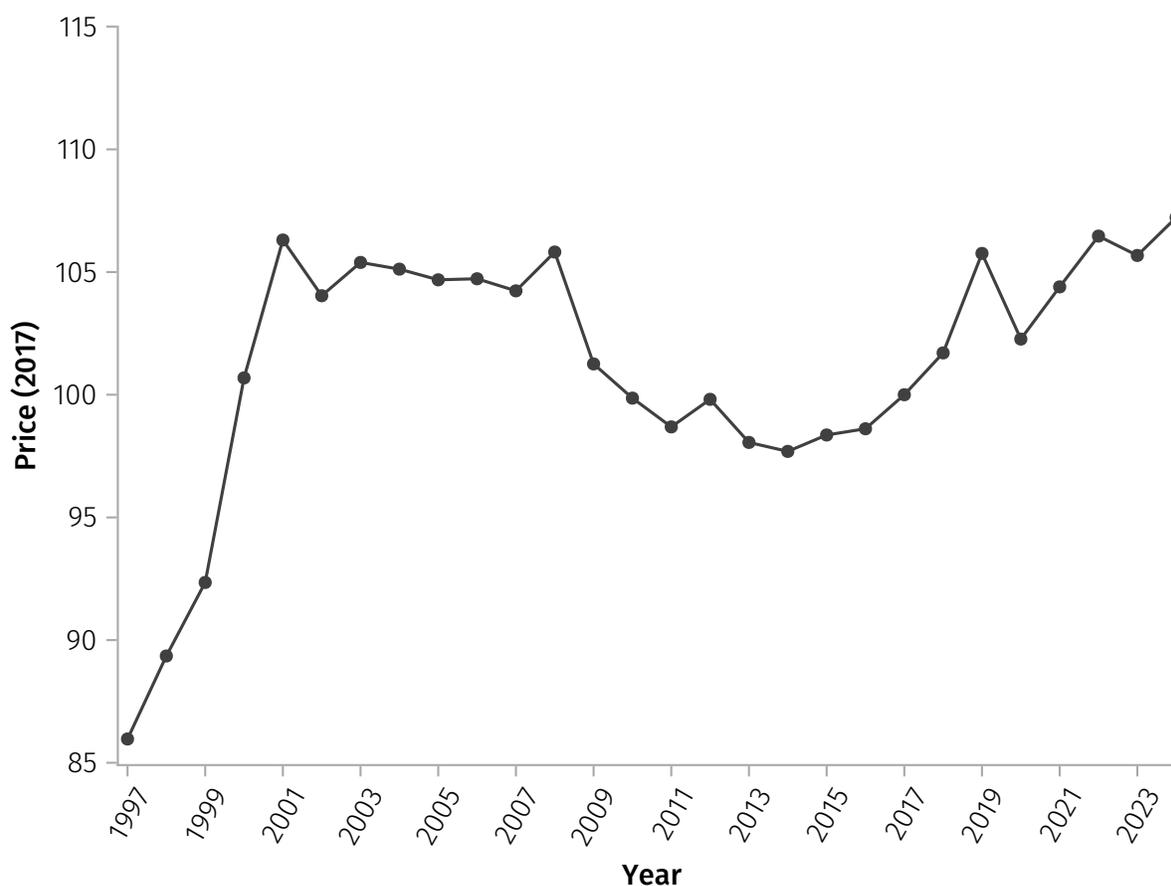
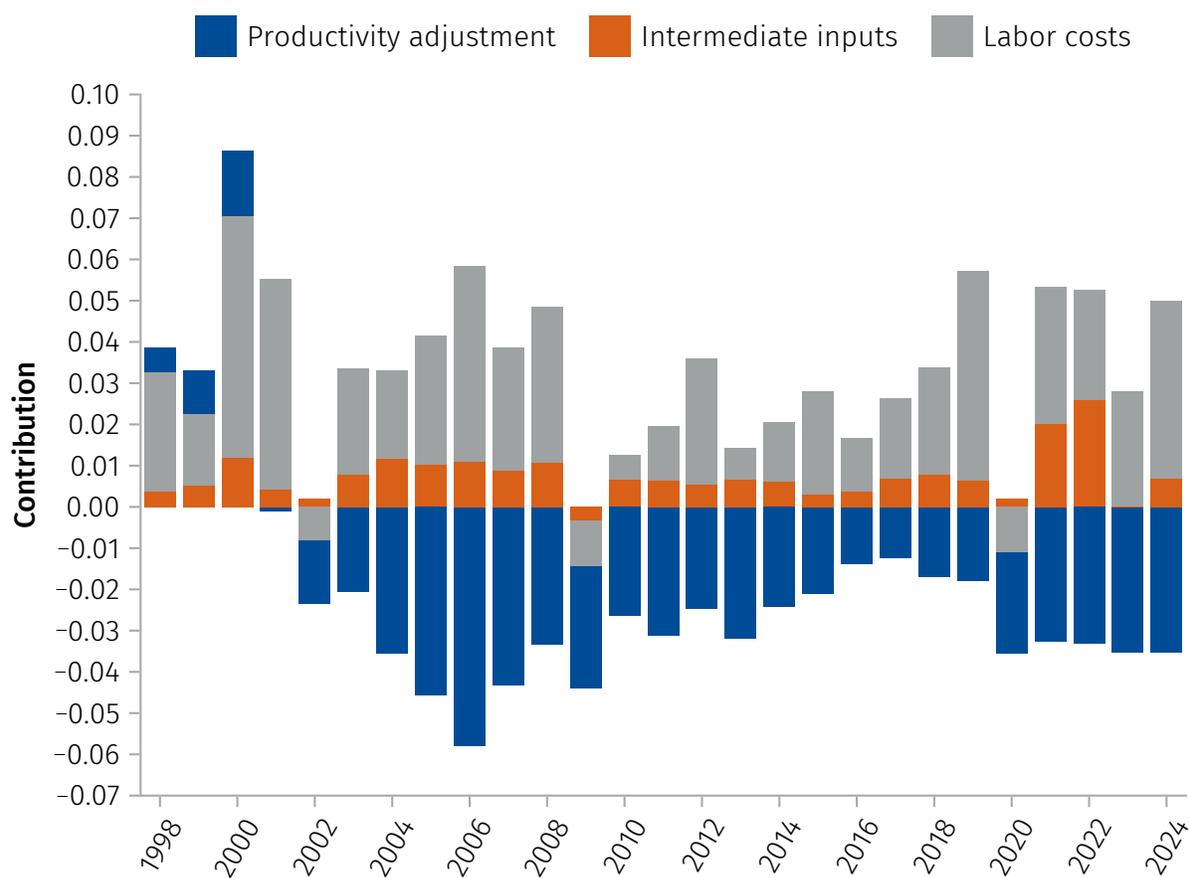


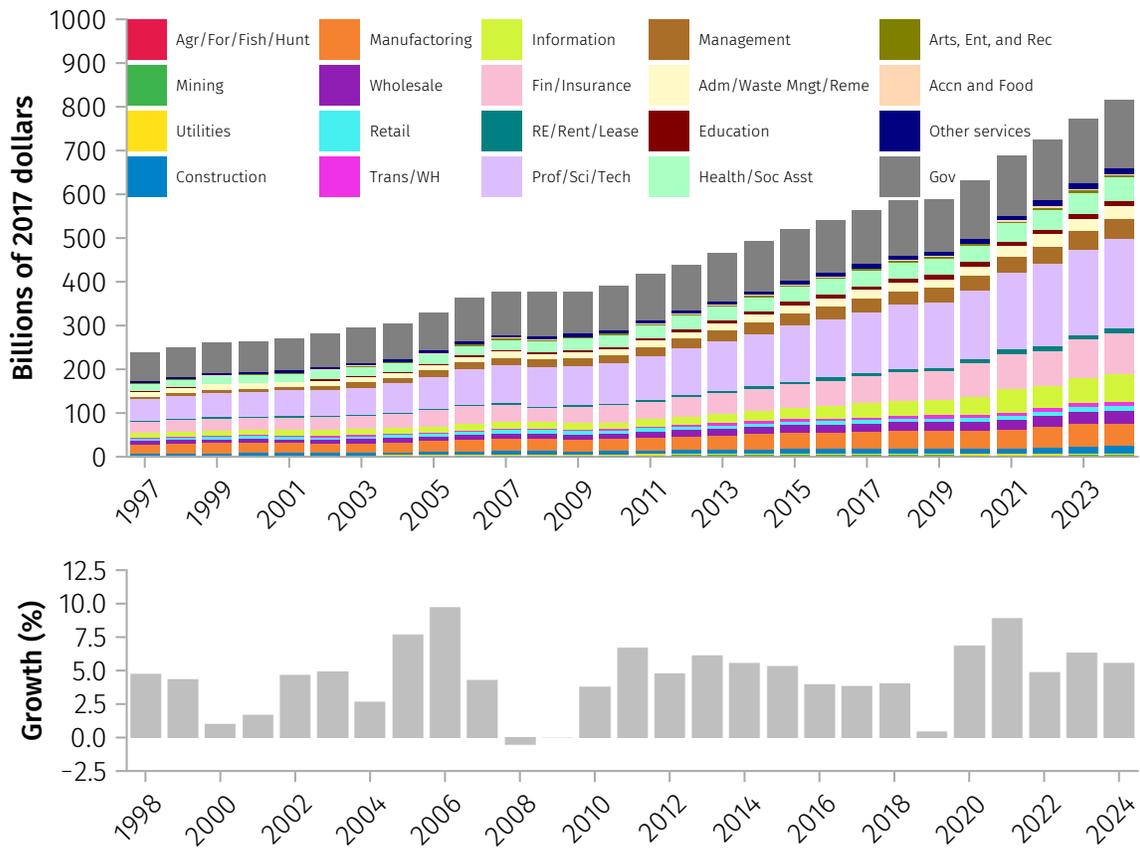
Figure 10. Composition of price movements



7.5. Volume

The real investment series, or volume, which is computed as the nominal investment deflated (divided) by the prices, shows GFCF ranging from \$239 billion in 1997 to \$816 billion in 2024 in 2017 dollars. Figure 11 shows the level and growth rates for the past two decades. The average log growth for the series was 4.5%.

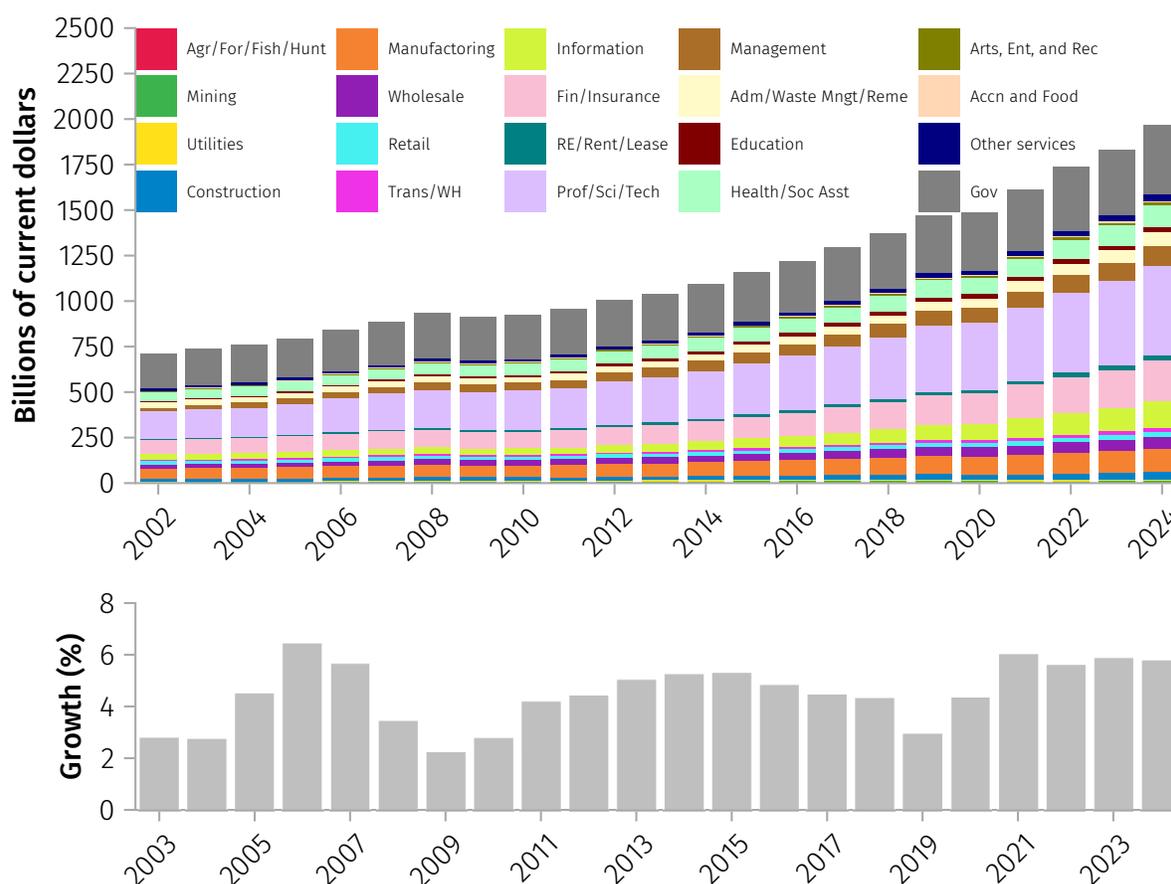
Figure 11. Volume



7.6. Net stock

The current-cost net stock series for 2002–2024 is shown in figure 12. Given the start of the series for the perpetual inventory method, we exclude the initial five years to account for one full average service life. The 1997 initialization seed was obtained by dividing the volume estimate by the depreciation rate (1/3). The series shows the current cost of the net stock for the asset class, which was \$710 billion in 2002 and increased to \$1,966 billion in 2024. The real average log growth for the series was 4.6%.

Figure 12. Current-Cost Net Stock

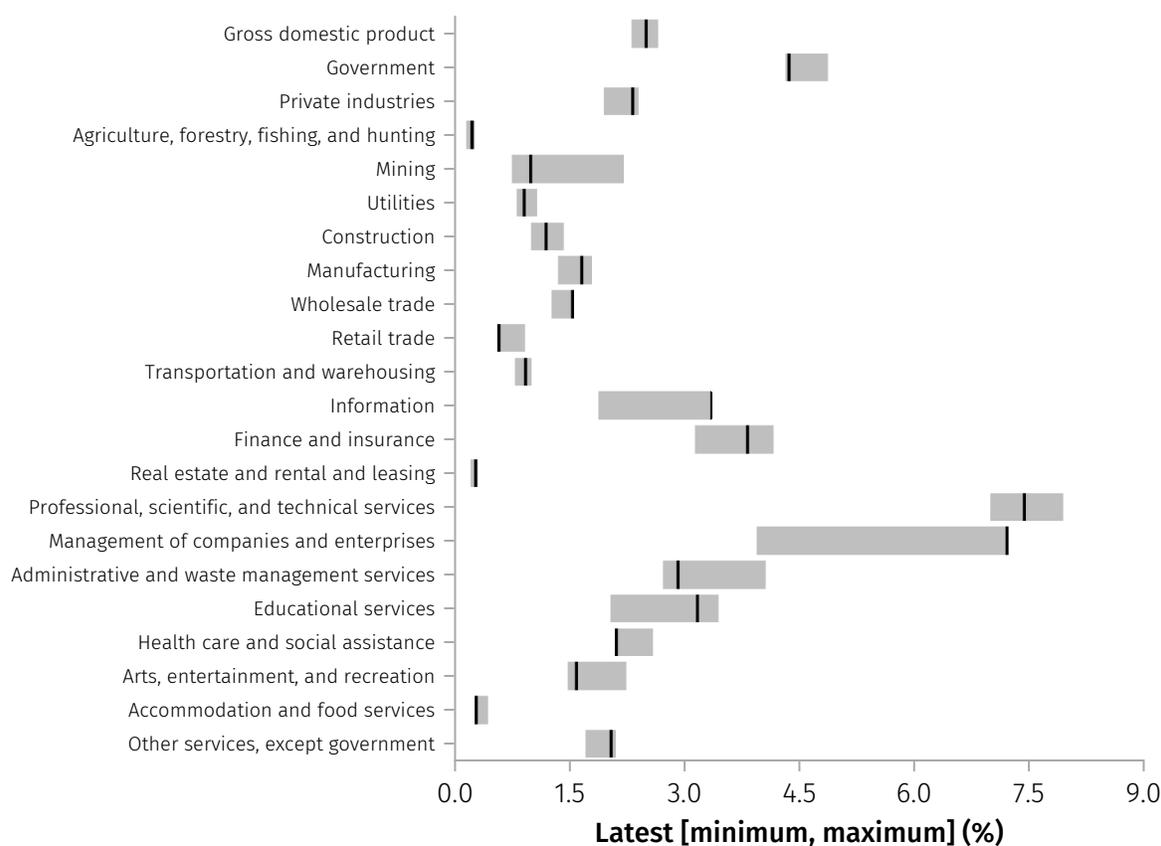


7.7. By Industry

In addition to considering the investment in the asset within each industry, we can assess how salient the asset is across industries using a measure of use, such as CFC as a share of industry size (i.e., value added). Figure 13 shows CFC as a share of value added for all industries (GDP), government (G), the private sector (PVT), and industry sectors. The statistics reported include the minimum, maximum, and latest values in the series from 2002–2024.

When accounting for industry size by value added, the industries with the highest relative use of data and database assets include PST (ILPA 54), government (ILPA: G), management of companies and enterprises (ILPA 55), and finance and insurance (ILPA 52). Most industries have increased their GFCF of the asset class during the last couple of decades, as shown in figure 13. Industries that have not increased investment in data and database assets relative to their size include agriculture, forestry, fishing, and hunting (ILPA 11); mining (ILPA 21); utilities (ILPA 22); retail and trade (ILPA 44RT); and transportation and warehousing (ILPA 48TW). Industries with relatively constant investment rates include construction (ILPA 53) and accommodation and food services (ILPA 72). The share of CFC to GDP (i.e., CFC / (Adjusted Published VA)) ranged from a low of 2.3% in 2006 to a high of 2.64% in 2020, with a start of 2.51% in 2002 and a most recent value of 2.54% in 2024.

Figure 13. Share of Consumption of Fixed Capital Over Value-Added (2002–2024)



8. Impact Across Accounts

This section examines the impact of capitalizing data across accounts, both in terms of the shares of investment in intellectual property products, nonresidential fixed assets, and government fixed assets, and in terms of its effects on value added and growth. The impact on growth is computed by adjusting the

nominal series and deflating it using the new adjusted Fisher prices, obtained by aggregating published prices and prices for data and database assets, with the corresponding weights.

8.1. Fixed assets

Investment in data and database (DD) assets for the private sector has ranged from 11.5% to 16% of nonresidential fixed assets. It has decreased slightly in the share of IPPs, from around 36% to 32%, with values at the lower end in recent years. The shares are computed by dividing the nominal estimates for GFCF in data assets by the sum of the published series and the nominal estimates for GFCF in data. Regarding government investment, the share of nonresidential fixed assets has ranged from 13% to 16%, while the share of IPPs has decreased from 36% in 1997 to 32% in 2024. The movement in the series is shown in figure 14. Some considerations can explain the trends. For example, nonresidential structures account for a significantly larger share of nonresidential fixed assets than in the private sector, thereby reducing investment in other asset categories. Moreover, government software investment has increased quite fast, as shown in figure 15.

In the private sector, investment in data and database assets is of the same order of magnitude as investment in R&D and software. Among the three categories, investment in software, excluding R&D software, has grown the fastest. One reason could be that the complementarity among asset classes boosts demand for software investment. For example, in the era of artificial intelligence (AI), including big data and generative AI, access to data and databases drives increased demand for software.

The volume series impacts the growth patterns of investment in fixed assets. Table 4 shows the impact of real growth investment on private nonresidential fixed assets, private IPPs, government IPPs, and government fixed assets. GFCF in data and databases grew more slowly than other IPP asset classes, such as software, in both the private and public sectors. However, the investment grew more rapidly than private nonresidential fixed assets and government fixed assets.

The changes in the net stock composition of government fixed assets, private nonresidential, and IPPs are shown in figure 16. Regarding the share of DD, it has decreased from 21.7% to 18.3% among government IPPs, from 26.4% to 22.6% among private IPPs, from 2.69% to 1.77% among government fixed assets, and remained around 4% for private non-residential fixed assets until 2018, with a recent increase up to 4.23% in 2024.

Figure 14. Share of IPPs and Nonresidential Fixed Assets GFCF

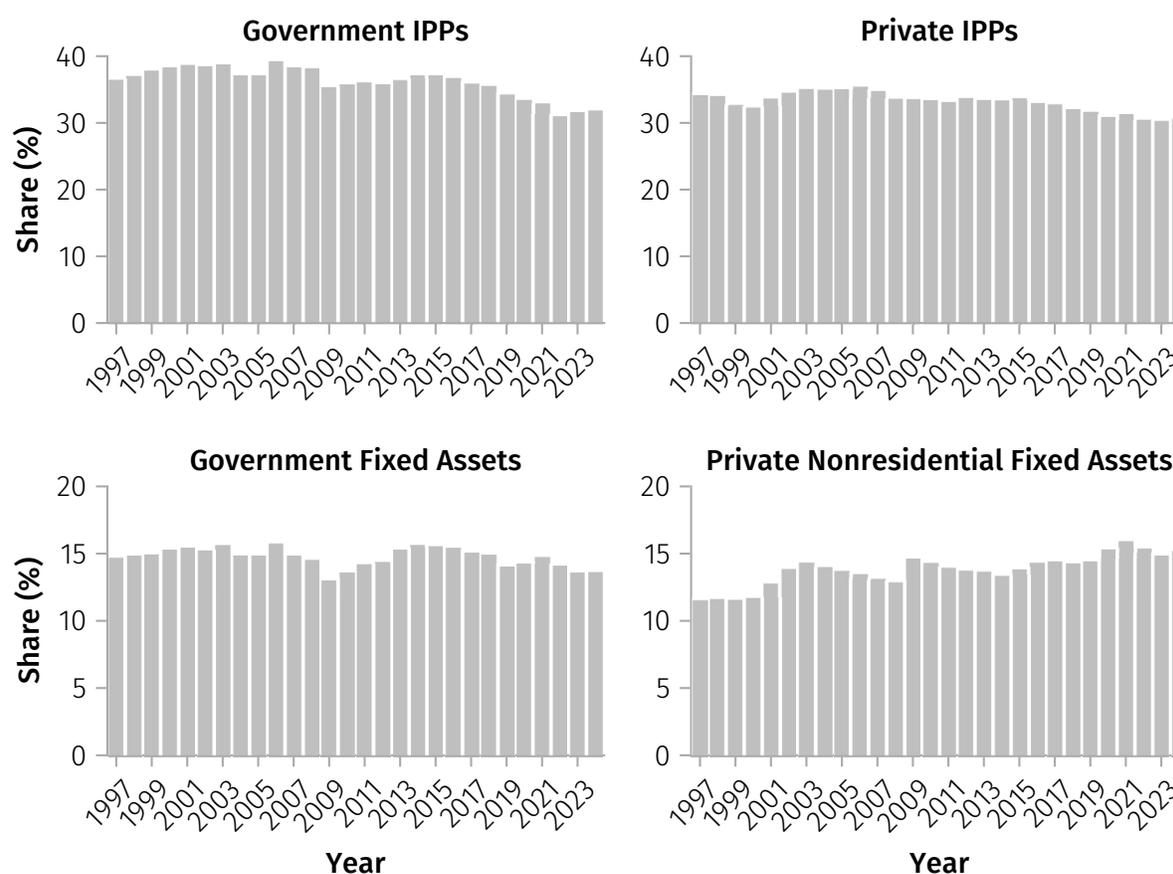


Table 4. Gross fixed capital formation compound annual growth rate with and without investment in own-account data and databases by aggregate 1997–2024 (%)

Account	Adjusted	Published	Δ
Pvt IPPs	5.61	5.85	-0.24
Pvt NFA	4.04	3.80	0.24
Gov IPPs	3.41	3.47	-0.06
Gov FA	2.16	2.00	0.16

Figure 15. Share of IPPs

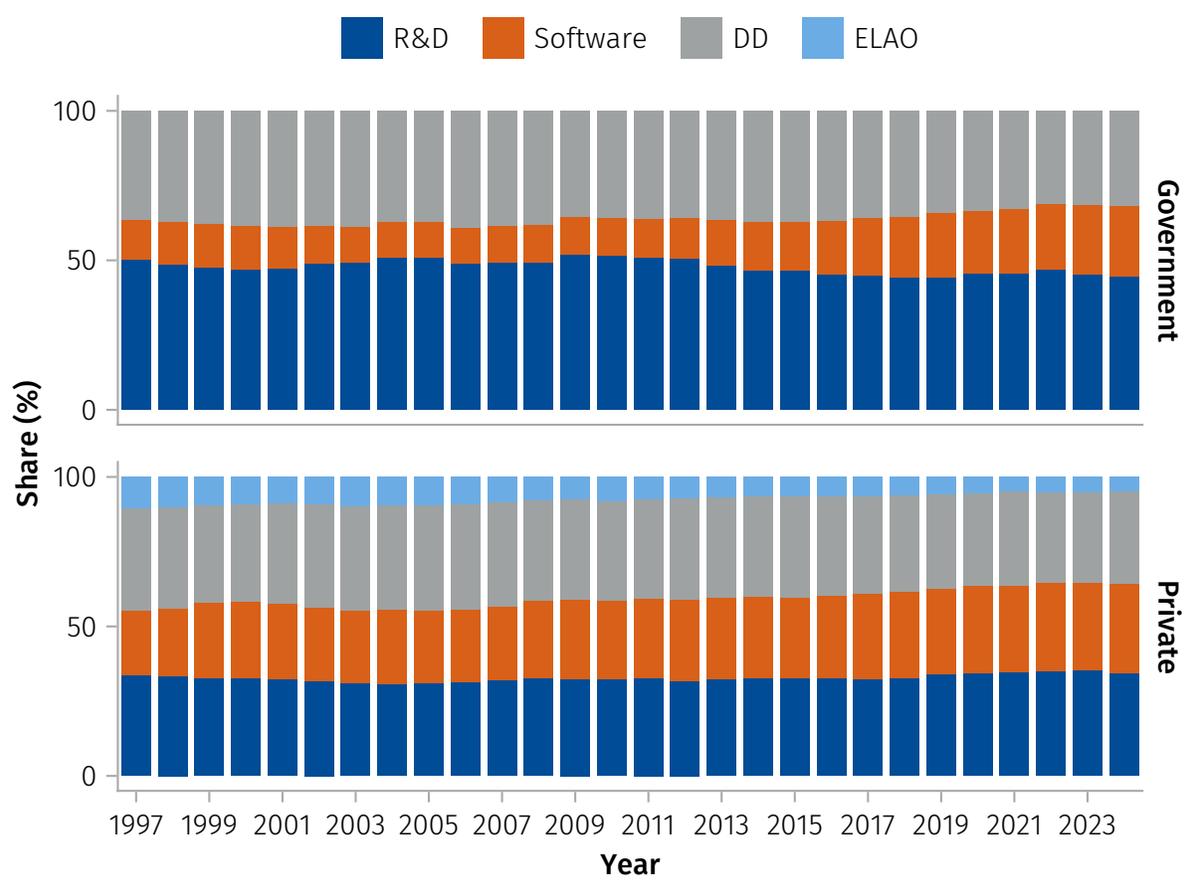
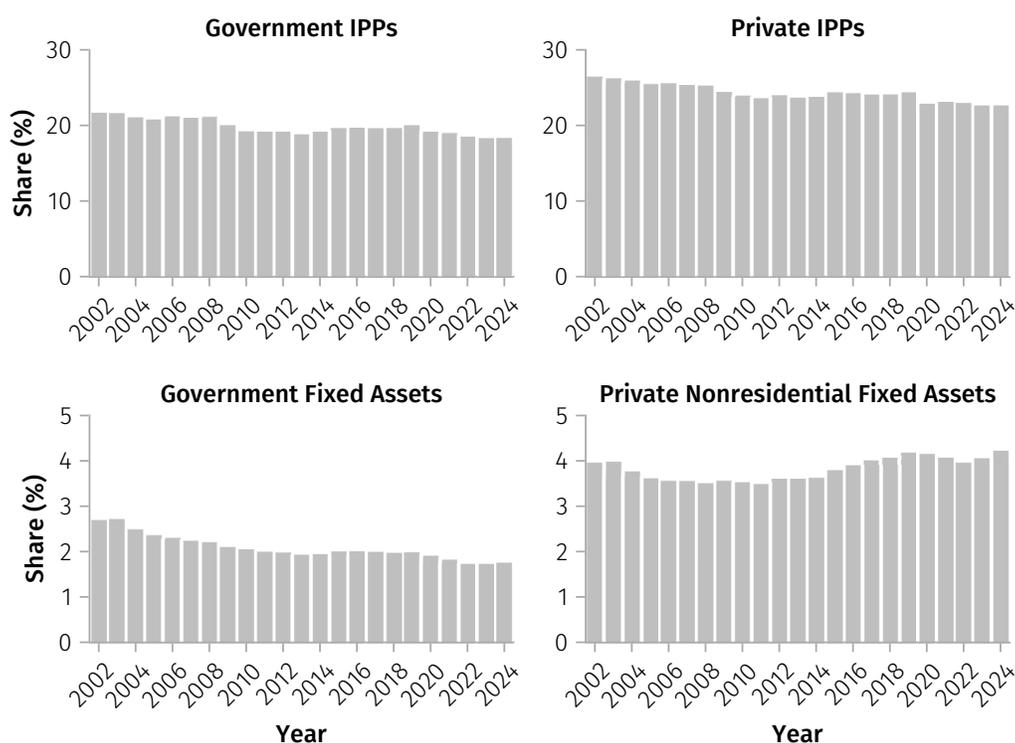


Figure 16. Share of Net Stock



8.2. Effect on value-added

The capitalization of own-account data and databases is associated with an average increase of 7 basis points in annualized GDP growth from 2002 to 2024. Tables 5 and 6 show the distribution of the impact by industry. The largest effects are on the management of companies and enterprises (ILPA 55), government (ILPA G), educational services (ILPA 61), finance and insurance (ILPA 52), and professional, scientific, and technical services (ILPA 54). The mechanism by which value added is affected is through the GFCF of private industries, as the capitalization of the new asset represents dual production, treating asset formation as additional output during that period. In the case of government investment, the impact on value added arises from the additional CFC in each period (other factors of production have already been accounted for, since government output is valued at cost). The impact on GDP results from the combined effects of private and government investment through the described mechanisms. The addition to value added is accounted for using asset-class prices, and the remainder is calculated using published prices to obtain the new adjusted series, with the new aggregated price used to compute quantity movements. Given that the government, and consequently GDP, rely on CFC estimates, we report the impact on the years following the initial average service life under the perpetual inventory method for the net stocks.

Table 5. Effect of capitalization on value-added compound annual growth rate for the private sector for 1997–2024 (%)

Account	Adjusted	Published	Δ
Private industries	2.61	2.56	0.06
Agriculture, forestry, fishing, and hunting	2.05	2.04	0.01
Mining	2.61	2.61	0.00
Utilities	1.27	1.25	0.03
Construction	0.08	0.02	0.06
Manufacturing	2.02	2.00	0.02
Wholesale trade	1.93	1.89	0.05
Retail trade	2.44	2.44	0.00
Transportation and warehousing	1.85	1.81	0.03
Information	6.60	6.58	0.02
Finance and insurance	2.32	2.22	0.10
Real estate and rental and leasing	2.62	2.61	0.01
Professional, scientific, and technical services	4.15	4.06	0.09
Management of companies and enterprises	3.28	2.99	0.29
Administrative and waste management services	3.30	3.29	0.00
Educational services	2.03	1.89	0.14
Health care and social assistance	2.97	2.94	0.03
Arts, entertainment, and recreation	2.17	2.11	0.06
Accommodation and food services	1.67	1.67	0.01
Other services, except government	-0.24	-0.34	0.09

Table 6. Effect of capitalization on value-added compound annual growth rate for the economy and government for 2002–2024 (%)

Account	Adjusted	Published	Δ
GDP	2.21	2.14	0.07
Government	0.68	0.57	0.11

9. Conclusion

This paper presents BEA's proposed implementation and estimates for capitalizing own-account data and databases as part of plans to transition to the newly adopted *System of National Accounts* standard. This experimental economic statistics previews the nominal investment, price, and volume for GFCF, estimates for CFC, and net stock annual estimates by industry for own-account data and database assets, as well as the impact across the accounts. Important ongoing discussions include the methodology choices for the asset boundary, occupation involvement rates, and the decisions made regarding pseudo-output prices and depreciation profiles. Relevant files posted with this article include the tables with the estimates, as well as the reproducibility repository that implements the methodology at <https://github.com/us-bea/reproducibility-toward-data-asset>.

Future updates will extend the estimates to encompass the relevant portion of the historical series and describe the various methodological breaks to achieve a consistent series. Additional work will cover other research areas, such as estimating and updating capital services and the productivity effects of expanding the production boundary.

References

- Advisory Expert Group on National Accounts (Mar. 31, 2023). "DZ.6 Recording of data in the National Accounts". In: *22nd Meeting of the Advisory Expert Group on National Accounts*. Vol. 2008 SNA Update. Guidance Notes Endorsed. Inter-secretariat Working Group on National Accounts. United Nations. URL: <https://unstats.un.org/unsd/nationalaccount/snaupdate/dztt.asp> (cit. on p. 1).
- Coyle, Diane and Annabel Manley (Sept. 2024). "What is the value of data? A review of empirical methods". In: *Journal of Economic Surveys* 38.4, pp. 1317–1337. ISSN: 0950-0804, 1467-6419. DOI: [10.1111/joes.12585](https://doi.org/10.1111/joes.12585) (cit. on p. 1).
- Eurostat (2026). *Handbook on measuring data in the system of national accounts*. Manuals and guidelines. Publications Office of the European Union. ISBN: 9789268298848. DOI: [10.2785/2811750](https://doi.org/10.2785/2811750) (cit. on p. 1).
- Fort, Teresa and Shawn Klimek (2018). *The Effects of Industry Classification Changes on US Employment Composition*. Working Papers 28. U.S. Census Bureau, Center for Economic Studies. URL: <https://www.census.gov/library/working-papers/2018/adrm/ces-wp-18-28.html> (cit. on p. 5).
- Hopson, Amy (Aug. 19, 2021). "Mapping Employment Projections and O*NET data: a methodological overview". In: *Monthly Labor Review*. ISSN: 19374658. DOI: [10.21916/mlr.2021.18](https://doi.org/10.21916/mlr.2021.18) (cit. on p. 4).

- Intersecretariat Working Group on National Accounts (Mar. 7, 2025). *2025 System of National Accounts*. Pre edit version. United Nations. URL: https://unstats.un.org/unsd/nationalaccount/docs/2025_SNA_Pre-edit.pdf (cit. on p. 1).
- Lightcast (Aug. 2023). *The Lightcast Open Skills Taxonomy: Understanding a Fast-Changing Labor Market*. URL: <https://lightcast.io/products/data/our-taxonomies> (cit. on p. 8).
- (2024). *Lightcast Taxonomies*. URL: <https://lightcast.io/our-taxonomies> (cit. on p. 7).
- Mitchell, John, Daniel Ker, and Molly Leshner (Dec. 21, 2021). *Measuring the economic value of data*. Going Digital Toolkit Notes 20. DSTI/CDEP/GD(2021)2/FINAL JT03487841. DOI: [10.1787/f46b3691-en](https://doi.org/10.1787/f46b3691-en) (cit. on p. 2).
- Moylan, Carol E. and Sumiye Okubo (2020). *The Evolving Treatment of R&D in the U.S. National Economic Accounts*. U.S. Bureau of Economic Analysis. URL: <https://www.bea.gov/system/files/2020-04/the-evolving-treatment-of-rd-in-the-us-national-economic-accounts.pdf> (cit. on p. 1).
- National Center for Science and Engineering Statistics (n.d.). *National Survey of College Graduates (NSCG): Public Use Data Files*. URL: <https://nces.nsf.gov/explore-data/microdata/national-survey-college-graduates> (cit. on p. 8).
- O*NET Resource Center (May 2024). *O*NET 28.3 Database*. URL: https://www.onetcenter.org/db_releases.html (cit. on pp. 7, 8).
- Parker, Robert P and Bruce T Grimm (May 2000). *Recognition of Business and Government Expenditures for Software as Investment: Methodology and Quantitative Impacts, 1959-98*. Tech. rep. U.S. Bureau of Economic Analysis. URL: <https://www.bea.gov/research/papers/2000/recognition-business-and-government-expenditures-software-investment> (cit. on p. 1).
- Rassier, Dylan G., Robert J. Kornfeld, and Erich H. Strassner (2019). “Treatment of Data in National Accounts”. In: URL: <https://www.bea.gov/system/files/2019-05/Paper-on-Treatment-of-Data-BEA-ACM.pdf> (cit. on p. 1).
- Santiago Calderón, J.B. and Dylan G. Rassier (2025). “Valuing the US Data Economy Using Machine Learning and Online Job Postings”. In: *Technology, Productivity, and Economic Growth*. University of Chicago Press. Chap. 8. URL: <https://www.nber.org/chapters/c14739> (cit. on pp. 1, 8).
- Schmidt, Julia, Graham Pilgrim, and Annabelle Mourougane (Sept. 18, 2023). *What is the role of data in jobs in the United Kingdom, Canada, and the United States?: A natural language processing approach*. OECD Statistics Working Papers 2023/05. Series: OECD Statistics Working Papers Volume: 2023/05. DOI: [10.1787/fa65d29e-en](https://doi.org/10.1787/fa65d29e-en). URL: https://www.oecd-ilibrary.org/economics/what-is-the-role-of-data-in-jobs-in-the-united-kingdom-canada-and-the-united-states_fa65d29e-en (cit. on p. 1).
- Statistics Canada (July 10, 2019). *The value of data in Canada: Experimental estimates*. URL: <https://www150.statcan.gc.ca/n1/pub/13-605-x/2019001/article/00009-eng.htm> (cit. on p. 1).

U.S. Bureau of Economic Analysis (Sept. 29, 2022). *BEA/BLS Integrated Industry-level Production Account 1987-2021*. URL: <https://www.bea.gov/system/files/2023-09/BEA-BLS-industry-level-production-account-1987-2021.xlsx> (cit. on p. 16).

– (2023a). “Chapter 6: Private Fixed Investment”. In: *NIPA Handbook: Concepts and Methods of the U.S. National Income and Product Accounts*. URL: <https://www.bea.gov/resources/methodologies/nipa-handbook> (cit. on p. 8).

– (2023b). *Table 6.7D. Self-Employed Persons by Industry (A) 1998 – Present*. Frequency: Annual; Status: Discontinued (cit. on p. 6).

– (2023c). *Table 6.9D. Hours Worked by Full-Time and Part-Time Employees by Industry*. Frequency: Annual; Status: Discontinued (cit. on p. 6).

– (Apr. 25, 2025). *BEA/BLS Integrated Industry-level Production Account 1997-2023*. URL: <https://www.bea.gov/system/files/2025-04/BEA-BLS-industry-level-production-account-1997-2023.xlsx> (cit. on pp. 6, 16).

U.S. Bureau of Labor Statistics (Apr. 25, 2024a). *Detailed industries: hours and employment*. URL: <https://www.bls.gov/productivity/tables> (cit. on p. 6).

– (2024b). *Occupational Outlook Handbook*. URL: <https://www.bls.gov/ooh> (cit. on pp. 7, 8).

– (n.d.). *Occupational Employment Statistics: National industry-specific and by ownership*. URL: <https://www.bls.gov/oes/tables.htm> (cit. on pp. 4, 8).

United Nations (Sept. 4, 2009). *System of National Accounts 2008*. United Nations. ISBN: 9789211615227. DOI: [10.18356/4fa11624-en](https://doi.org/10.18356/4fa11624-en) (cit. on p. 1).