R&D Labor Shares

Paul J. Hamilton
Harvard University

Justine Boudou
London Business School

Todd Lensman
Harvard University

Kyle R. Myers
Harvard University & NBER

July 2026
PRELIMINARY AND INCOMPLETE

We assemble annual US data on R&D labor compensation from 1981 through 2022 and compare this compensation with reported R&D expenditures and model-estimated R&D-based income. Under our baseline assumptions, R&D-based income is roughly 13.0 percent of GDP, and R&D labor compensation averages 10.9 percent of this income. The income share increases from 8.1 percent in 1981 to 16.6 percent in 2022; an exact percentage-change decomposition shows that real compensation per measured worker and worker headcount raise the share, while real GDP growth offsets much of their combined contribution. Compensation also rises relative to reported accounting expenditures, but these ratios are not economic cost shares because the source accounts do not construct a common user cost of R&D capital.

Hamilton: phamilton@hbs.edu. Boudou: jboudou@hbs.edu. Lensman: tlensman@hbs.edu. Myers: kmyers@hbs.edu. This project received financial support from the Harvard Business School. There are no financial relationships or other potential conflicts of interest that apply to any of the authors.

1. Introduction

Labor shares are important. First, there is the *labor share of income*: the share of income produced by an economy, sector, or firm that is paid to workers for their labor services. Second, there is the *labor share of costs*: the share of the economic cost of production attributable to labor rather than other factors. A tremendous body of work has investigated these labor shares, but not as they occur in research and development (R&D). In this paper, we assemble and investigate new data on R&D labor shares in the US over the past forty years.

The labor share of income is a longstanding political economy question (Ricardo 1817; Karabarbounis 2024). In the case of R&D, the direct labor compensation share of R&D-based income is one margin of appropriability, with potential consequences for innovation and growth (Romer 1990; Aghion and Howitt 1992; Grossman and Helpman 1991; Jones 1995). Unlike the metric applied to final-goods labor, where income is an accounting object, the metric applied to R&D requires an economic model since R&D-based income cannot be directly observed. Perhaps due to this modeling requirement, this metric has received little, if any, attention.

The labor share of costs is a key ingredient for understanding production functions, the elasticity of substitution between factors, and technological change. Advances in automation have heightened interest in documenting and understanding trends in cost shares across the economy (Karabarbounis and Neiman 2014; Acemoglu and Restrepo 2018; Autor et al. 2020). Furthermore, emerging capabilities have led many to ask how the "idea production function" might evolve and whether R&D itself can be automated (Aghion, Jones, and Jones 2019; Jones 2025). However, factor-specific data on R&D have gone largely unassembled, with nearly all empirical investigations treating R&D investment as involving a single factor.

To shed new light on these objects, we assemble annual US data from 1981 through 2022 on R&D expenditures, labor compensation, and employment in business, higher education, government, and nonprofit institutions. We harmonize multiple public data sources to provide a new view of the evolution of US R&D labor.

The data reveal substantial growth in recorded R&D labor compensation. In 2022 dollars, total compensation increases from $93.0 billion in 1981 to $560.7 billion in 2022. Business performers account for most of this amount, with compensation increasing from $69.1 billion to $472.6 billion. Compensation in higher education, government, and nonprofit institutions increases from $23.8 billion to $88.1 billion.

The accounting cost shares reveal interesting trends as well. Business labor compen-

sation increases from 48.7 percent of reported expenditures in 1981 to 68.3 percent in 2022. The corresponding ratio for non-business R&D increases from 40.2 percent to 44.5 percent.

We estimate R&D-based income using the delay approach of Jones and Williams (1998) and Jones and Summers (2022). R&D-based income is defined as the present value of aggregate output lost if the productivity improvement associated with a year's idea flow arrived one year later. Under our baseline assumptions, estimated R&D-based income equals 13.0 percent of contemporaneous GDP. Labor compensation averages 10.9 percent of this income and increases from 8.1 percent in 1981 to 16.6 percent in 2022. Alternative discount rates, horizons, and trend-growth measures affect the level of the income share, although most specifications reveal doubling over the sample period.

Most of the increase reflects compensation paid by business performers. Business compensation's contribution to the economy-wide share increases from 6.1 percent in 1981 to 14.0 percent in 2022. The corresponding contribution from non-business R&D changes from 2.1 percent to 2.6 percent.

What accounts for the increase in R&D labor's share of income? We use a statistical decomposition to separate the three forces: compensation per R&D worker, R&D-worker headcount, and real GDP. Between 1981 and 2021, compensation per worker contributes 132.56 percentage points and headcount contributes 166.79 percentage points, while real GDP contributes −200.23 percentage points. Together, these terms produce a 99.12 percent increase in the income share; the doubling. The negative GDP contribution follows mechanically from the valuation approach: a larger output base raises the estimated income associated with a given proportional productivity improvement, so compensation must keep pace with GDP for labor's share to remain constant.

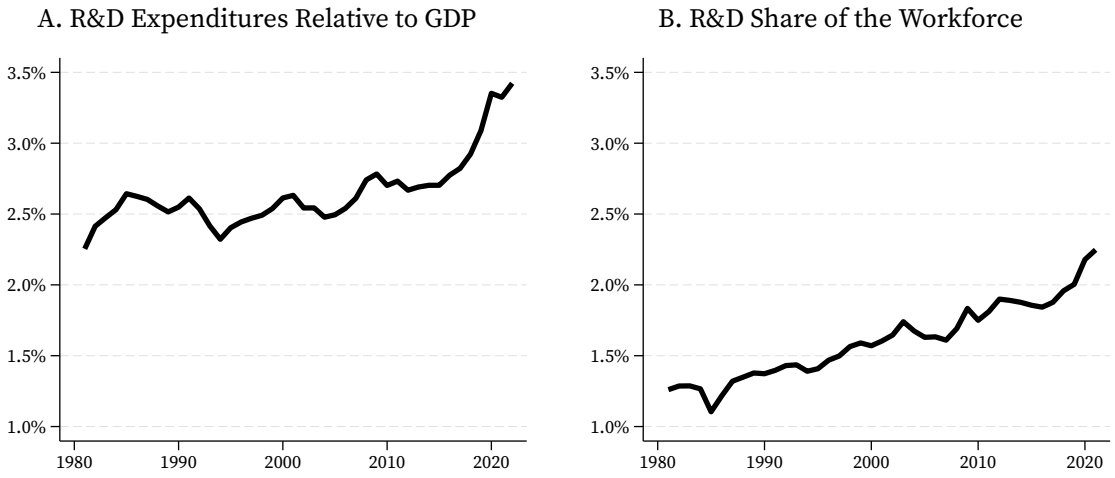
The paper proceeds as follows. Section 2 describes the data and the descriptive compensation facts. Section 3 estimates R&D-based income and reports the share paid to R&D labor. Section 4 turns to R&D labor's cost share [INCOMPLETE]. The final section discusses the economic implications.

Related Literature

Endogenous-growth models place research effort at the center of idea production and long-run growth (Romer 1990; Aghion and Howitt 1992; Jones 1995). Figure 1 places the growth of that effort in the context of the US economy. Between the endpoints of our sample, R&D expenditures increase from 2.25 percent to 3.42 percent of GDP, while the R&D share of the workforce increases from 1.26 percent to 2.25 percent through 2021. This expansion of research inputs has not been matched by a corresponding acceleration in

productivity growth. Jones (2009) relates this pattern to the growing burden of knowledge, and Bloom et al. (2020) documents broad declines in research productivity.

FIGURE 1
Aggregate R&D Activity



Note: Panel A divides national R&D expenditures by nominal GDP. Panel B divides a constructed, non-quality-adjusted R&D-worker headcount by OECD total employment; the worker series ends in 2021.

Our paper also relates to research on labor's share of income. This literature documents large changes in aggregate labor shares and studies the roles of capital prices, firm reallocation, and automation (Karabarbounis and Neiman 2014; Autor et al. 2020; Acemoglu and Restrepo 2018; Karabarbounis 2024). Our data provides a way to begin similar analyses of R&D.

The compensation numerator is shaped by the labor market for R&D workers. Government R&D demand can bid up the wages of scientists and engineers when specialized labor supply is inelastic (Goolsbee 1998). Over longer horizons, engineering enrollment and occupational choice respond to expected career prospects (Ryoo and Rosen 2004). Scientists may also accept lower pay for jobs that permit greater engagement with science (Stern 2004). At the same time, workers can capture part of the rents created by innovation (Van Reenen 1996; Kline et al. 2019). These mechanisms give different interpretations to compensation growth. We cannot yet isolate the causes of the trends we identify.

2. Data & Descriptive Statistics

2.1. Data Sources

Our main sample covers the United States from 1981 through 2022. We assign R&D to the institution that performs it, so the sector classification matches the workers in the

compensation numerator. We distinguish four performing sectors: business, higher education, government, and nonprofit institutions. Business R&D is further divided into industries.

The expenditure data come primarily from federal surveys of R&D performers. For business R&D, we join the successive NCSES–Census surveys (i.e., SIRD, BRDIS, BRDS, and BERD). Higher-education data combine the Higher Education R&D Survey with IPEDS. Government and nonprofit data combine NSF National Patterns with the BEA R&D Satellite Account. The worker series converts OECD researcher full-time equivalents to a broader R&D-worker headcount using overlap-period ratios from the four performing sectors; it ends in 2021. The income calculation adds BEA output and price data and total factor productivity from Penn World Table 11.

We define labor compensation to include payments for labor services, not only salaries and wages. Business compensation includes salaries, wages, employer benefits, temporary staffing, and stock-based compensation. Higher-education, government, and nonprofit compensation includes salaries, wages, and fringe benefits. Temporary-staffing expenditures may include agency margins and need not be paid entirely to workers. We use the BEA GDP price deflator to convert every component to 2022 dollars.

Coverage and measurement quality vary over time. The detailed business cost module is available in 34 survey years, and survey redesigns change which compensation components are reported separately. We manually harmonize industries across the SIC–NAICS transition; much of the detailed pre-1999 nonmanufacturing panel and selected manufacturing series are projected. Higher-education compensation is backcast for 1981–1986, and four later years fail the direct-data quality screen. Government and nonprofit compensation are observed from 2012 onward; for earlier years, we apply each sector's average 2012–2023 compensation-to-expenditure ratio to its historical expenditures. Because the data include interpolations, projections, and a constructed worker series, we emphasize broad trends rather than isolated annual changes, especially for early non-business and narrow-industry comparisons.

Appendix A describes the sources and harmonization in detail.

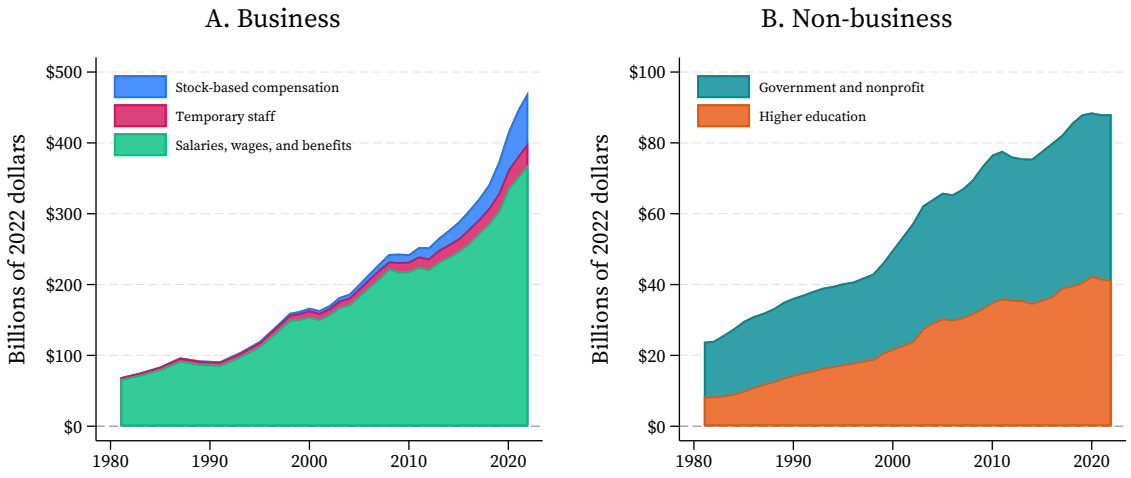
2.2. Trends in Labor Expenditures

Figure 2 reports R&D labor compensation by performing sector in 2022 dollars. Panel A separates business compensation into salaries, wages, and benefits; temporary staffing; and stock-based compensation. Panel B separates higher education from government and nonprofit performers.

Business compensation rises from $69.1 billion in 1981 to $472.6 billion in 2022. Over the same period, non-business compensation rises from $23.8 billion to $88.1 billion. Within the latter total, higher-education compensation rises from $8.4 billion to $41.3 billion, while government and nonprofit compensation rises from $15.5 billion to $46.8 billion. Business R&D thus accounts for most of both the level and growth of labor compensation.

In 2022, salaries, wages, and benefits account for 78.3 percent of business labor compensation, stock-based compensation for 15.2 percent, and temporary staffing for 6.6 percent.

FIGURE 2
Total R&D Labor Compensation by Performing Sector



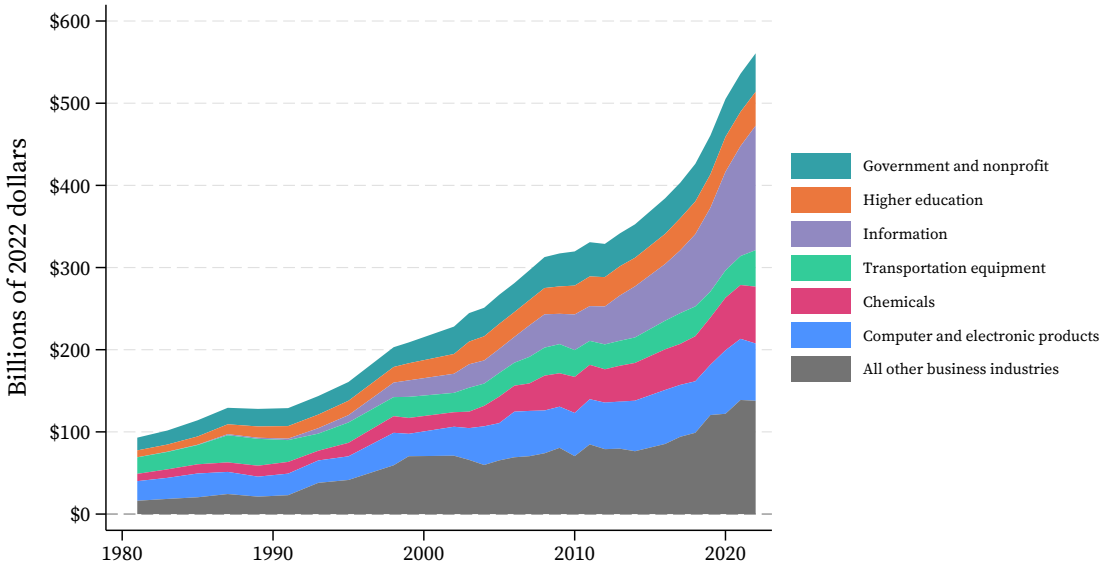
Note: Stacked areas report billions of 2022 dollars, using the GDP price deflator. Components are imputed when not separately reported; see Appendix A.

Figure 3 shows how R&D labor compensation is distributed across mutually exclusive and exhaustive industries and sectors. We select the four business industries with the largest cumulative real R&D expenditures over the full sample and hold that membership fixed: Computer and electronic products, Chemicals, Transportation equipment, and Information. The remaining business industries form one group. Higher education forms another group, and government and nonprofit R&D form the final group.

The industry composition shifts markedly. By 2022, information accounts for $151.5 billion in R&D labor compensation, compared with $69.7 billion in computer and electronic products, $69.2 billion in chemicals, and $44.4 billion in transportation equipment. All other business industries together account for $137.8 billion. Compensation growth is broad but increasingly concentrated in information.

Figure 4 divides R&D labor compensation by the expenditure total reported for the same sector. This is an accounting ratio, not an economic cost share, because the reported

FIGURE 3
Total R&D Labor Compensation by Business Industry and Non-business Sector



Note: Values are billions of 2022 dollars, using the GDP price deflator. The four named business industries have the largest cumulative real R&D expenditures over the sample and are held fixed over time. The pre-1999 industry allocation is partly projected, especially in nonmanufacturing, but leaves each year's published business total unchanged.

expenditure measures do not use a common user cost for R&D capital. For business R&D, the ratio rises from 48.7 percent in 1981 to 68.3 percent in 2022. For non-business R&D, it rises from 40.2 percent to 44.5 percent.[1]

3. R&D Labor's Share of Income

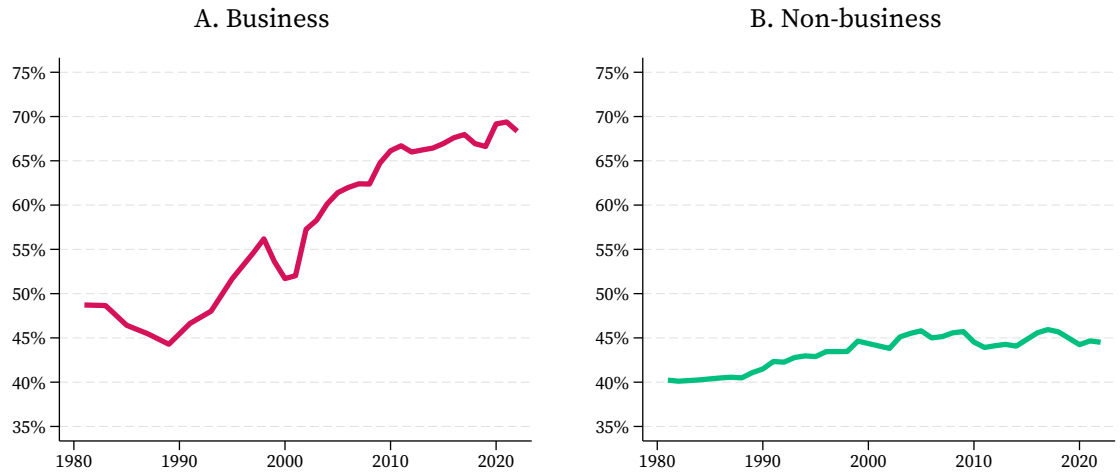
3.1. Valuation Framework

We measure R&D labor's share of income by comparing labor compensation with R&D-based income. Unlike the denominator in a conventional factor-income share, R&D-based income is not recorded contemporaneously in the national accounts. It can only be defined with a model. We define it as the present value of aggregate output associated with a year's idea flow.

Let W_t denote nominal R&D labor compensation and let V_t denote nominal estimated

[1]Before 2012, the government and nonprofit series use constant sector-specific compensation-to-expenditure ratios. Higher-education compensation is backcast before 1987. These construction choices partly account for the stability of the non-business series.

FIGURE 4
R&D Labor Compensation as a Share of Accounting Expenditures



Note: Each panel divides R&D labor compensation by the source-reported expenditure measure for the same sector. Non-business combines higher education, government, and nonprofit performers. The source accounts do not construct a common user cost of R&D capital, so these ratios are not economic cost shares.

R&D-based income. R&D labor's share of income is

$$s_{Lt}^{I} = \frac{W_t}{V_t}. \tag{1}$$

We match compensation recorded in year t to the value assigned to year t's idea flow.

We estimate R&D-based income V_t using the delay approach of Jones and Williams (1998) and Jones and Summers (2022). Under the maintained attribution of measured TFP growth to R&D, consider the productivity improvement generated by R&D performed in year t. We compare the observed path with a counterfactual in which the aggregate productivity path remains one year behind. R&D-based income is the present value of the resulting output gap.

Let Y_t denote real final output, D_t the GDP price deflator normalized so that $D_{2022} = 1$, g_A the constant annual growth rate of total factor productivity (TFP) in the final output sector, g_Y the annual growth rate of real output, r the discount rate, and H the number of future years considered. Delaying a TFP improvement of g_A by one year lowers the level of TFP, relative to its observed path, by

$$\delta_A = 1 - \frac{1}{1 + g_A}. \tag{2}$$

The real value of the delayed TFP improvement over horizon H is

$$\begin{aligned}\widetilde{V}_t(H) &= \delta_A \times Y_t \times \sum_{k=1}^{H} \left(\frac{1+g_Y}{1+r}\right)^k \\ &= Y_t \times \frac{\delta_A \times (1+g_Y)}{r-g_Y} \times \left[1 - \left(\frac{1+g_Y}{1+r}\right)^H\right]. \end{aligned} \tag{3}$$

For an infinite horizon and $r > g_Y$, this becomes

$$\widetilde{V}_t = \kappa \times Y_t, \qquad \kappa \equiv \frac{\delta_A \times (1+g_Y)}{r-g_Y}. \tag{4}$$

The product $\delta_A \times Y_t$ scales the output loss, and the geometric sum projects and discounts that loss over future years. The nominal income denominator in Equation (1) is $V_t = D_t \times \widetilde{V}_t = \kappa \times D_t \times Y_t$. Thus, the numerator and denominator can equivalently be compared in nominal dollars or in 2022 dollars.

The calculation attributes the measured trend in aggregate TFP to the annual R&D idea flow and treats the resulting TFP gap as a proportional output gap, holding other inputs and behavioral responses on their observed paths. The baseline uses annualized growth rates implied by the 1981 and 2022 endpoints for every sample year. It is therefore a retrospective valuation, not a measure of expectations held in year t. Annual TFP growth is 0.61 percent and annual real GDP growth is 2.69 percent. With a 7.5 percent real discount rate and an infinite horizon, the estimated present value of one year's idea flow is 13.00 percent of contemporaneous GDP, so $\widetilde{V}_t/Y_t = \kappa = 13.00$ percent.

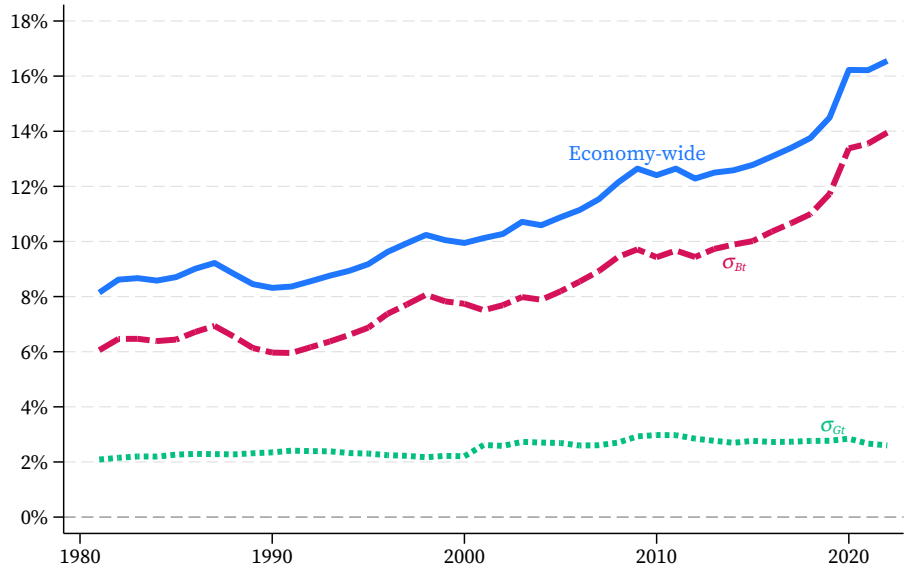
3.2. Trends in R&D Labor's Share of Income

Figure 5 reports the baseline estimates. R&D labor's share of income averages 10.9 percent over the sample. It rises from 8.1 percent in 1981 to 16.6 percent in 2022, with most of the increase occurring in the later years. Under the baseline, κ is fixed. The time path is therefore a constant rescaling of R&D labor compensation relative to GDP; it does not use annual movements in TFP to produce annual movements in the share.

Which performing sector's labor compensation accounts for the aggregate ratio? We split the numerator between business and non-business performers while retaining the common economy-wide denominator. Let W_{Bt} denote business R&D labor compensation and W_{Gt} compensation in higher education, government, and nonprofit R&D. Define the following contributions:

$$\sigma_{Bt} = \frac{W_{Bt}}{V_t}, \qquad \sigma_{Gt} = \frac{W_{Gt}}{V_t}, \qquad \sigma_{Bt} + \sigma_{Gt} = s^I_{Lt}. \tag{5}$$

FIGURE 5
R&D Labor's Share of Income



Note: R&D-based income is the present value of output lost if the TFP path is delayed by one year. The baseline uses annualized 1981–2022 endpoint growth, a 7.5 percent real discount rate, and an infinite horizon. Business and non-business lines split only the compensation numerator and use the same economy-wide denominator; they are not sector-specific income shares. Non-business combines higher education, government, and nonprofit performers. In 12 flagged years, an unavailable sector is completed as the residual from aggregate compensation.

Figure 5 shows trends in these contributions. The business contribution rises from 6.1 percent in 1981 to 14.0 percent in 2022. The non-business contribution changes much less, from 2.1 percent to 2.6 percent.

These differences do not necessarily reflect the productivity of business and non-business R&D. Suppose, for interpretation, that aggregate R&D-based income can be allocated additively across performing sectors. Let $\omega_{jt} = V_{jt}/V_t$ be sector j's unknown share of aggregate R&D-based income and let $\lambda_{jt} = W_{jt}/V_{jt}$ be its labor share of R&D-based income. The plotted contribution then satisfies

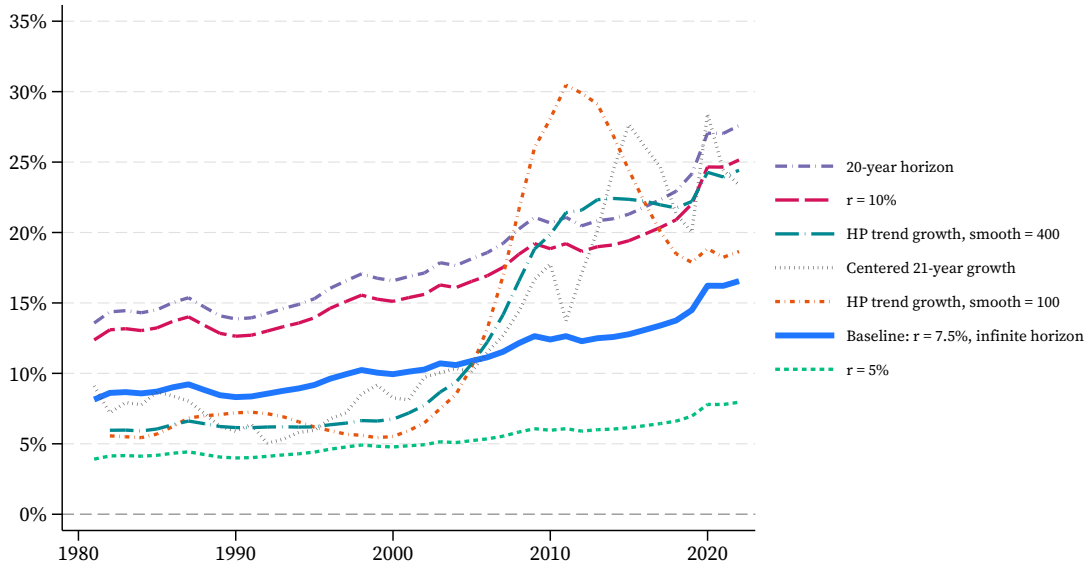
$$\sigma_{jt} = \omega_{jt} \times \lambda_{jt}. \tag{6}$$

The data identify this product, not either component. The distinction is particularly important for non-business R&D. Models of public research emphasize downstream spillovers (Akcigit, Hanley, and Serrano-Velarde 2021), and empirical work links public and academic research to later private innovation and aggregate productivity (Fleming

et al. 2019; Myers and Lanahan 2022; Fieldhouse and Mertens 2023). A small non-business contribution directly says only that non-business labor compensation is small relative to aggregate R&D-based income. Under an additive value allocation, this can reflect a smaller share of aggregate value, a smaller fraction of value paid directly to current labor, or both.

Figure 6 reports the results based on estimating R&D-based income under alternative assumptions. Lowering the discount rate to 5 percent increases the present value of future output and lowers the average labor share to 5.3 percent. Raising the rate to 10 percent lowers R&D-based income and raises the average labor share to 16.6 percent. Truncating the income stream after 20 years raises the average share to 18.2 percent. Specifications using centered growth rates or HP-filtered trends allow κ to change over time and produce more variation.

FIGURE 6
Robustness of R&D Labor's Economy-Wide Share of Income



Note: All series report $100 \times W_t/V_t$. The baseline uses annualized endpoint growth, a 7.5 percent real discount rate, and an infinite horizon. Alternatives vary the discount rate, horizon, or growth-rate specification. The local specification uses a centered 21-year window in interior years and truncated windows near the endpoints; HP trends are two-sided ex post filters.

The level of the income share depends materially on the discount rate, horizon, and maintained attribution of TFP growth. All specifications end above their early-sample values, but the level, timing, and sign of shorter-run movements depend on how trend growth is estimated. The local-growth specification uses a centered window in interior

years and truncated windows near the endpoints. The HP specifications are two-sided ex post filters and are especially sensitive near the endpoints.

In the baseline, the ratio averages 10.9 percent and increases from 8.1 percent to 16.6 percent. Its level is highly sensitive to the valuation assumptions. The increase means that recorded R&D labor compensation grew relative to estimated R&D-based income; it does not by itself identify a change in workers' appropriability.

3.3. Prices and Quantities

With κ fixed, three objects determine the income share: real compensation per measured R&D worker, R&D-worker headcount, and real GDP. Growth in the first two raises the share, while growth in real GDP lowers it. Let N_t denote measured R&D-worker headcount and define real compensation per measured worker as $P_t = (W_t/N_t)/D_t$. The baseline income share is

$$s_{Lt}^{I} = \frac{W_t}{\kappa \times D_t \times Y_t} = \frac{1}{\kappa} \times P_t \times N_t \times \frac{1}{Y_t}. \tag{7}$$

The compensation term P_t includes wages, benefits, temporary-staffing expenditures, stock-based compensation, and changes in the composition of measured workers. It is therefore not an average wage for a fixed-quality worker. The normalization of D_t fixes the units of P_t and Y_t but does not affect their growth rates.

Panel A of Figure 7 reports the ordinary percentage change in each object relative to 1981. To decompose the income share in the same units, define $p_t = P_t/P_0$, $n_t = N_t/N_0$, and $z_t = Y_0/Y_t$. Then

$$100 \times \left(\frac{s_{Lt}^{I}}{s_{L0}^{I}} - 1 \right) = 100 \times (p_t \times n_t \times z_t - 1). \tag{8}$$

Because the three factors multiply, their ordinary percentage changes contain interaction terms. We allocate those interactions symmetrically with a Shapley decomposition. For each factor, the method averages its marginal contribution over the six possible orders in which p_t, n_t, and z_t can move from one to their observed values. The resulting percentage-point contributions are

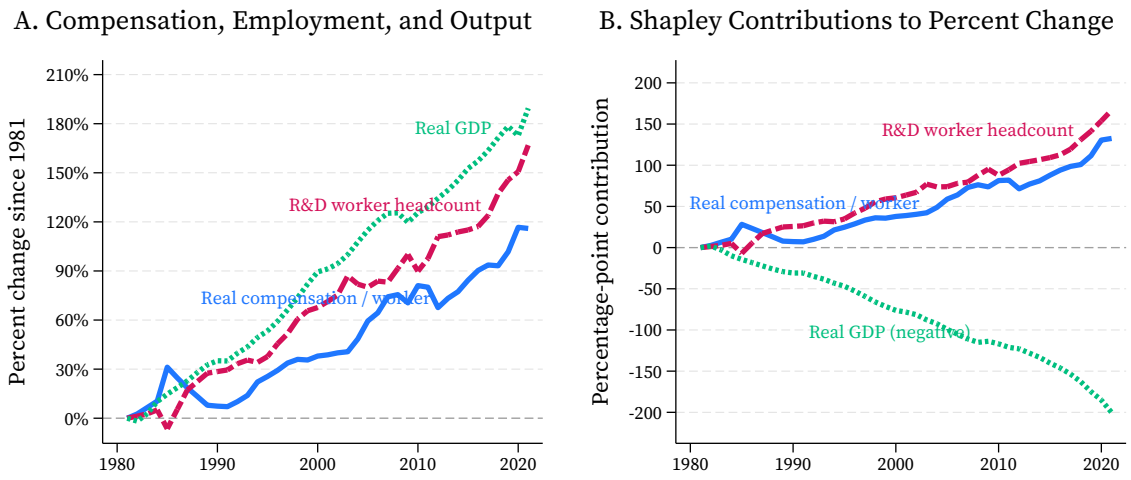
$$\begin{aligned}
C_t^P &= 100 \times (p_t - 1) \times \left[\frac{1}{3} + \frac{n_t + z_t}{6} + \frac{n_t \times z_t}{3} \right], \\
C_t^N &= 100 \times (n_t - 1) \times \left[\frac{1}{3} + \frac{p_t + z_t}{6} + \frac{p_t \times z_t}{3} \right], \\
C_t^Y &= 100 \times (z_t - 1) \times \left[\frac{1}{3} + \frac{p_t + n_t}{6} + \frac{p_t \times n_t}{3} \right].
\end{aligned} \tag{9}$$

These contributions satisfy

$$100 \times \left(\frac{s^I_{Lt}}{s^I_{L0}} - 1 \right) = C^P_t + C^N_t + C^Y_t . \tag{10}$$

Panel B plots the three contributions.

FIGURE 7
Decomposing Growth in R&D Labor's Share of Income



Note: Panel A reports percentage changes since 1981 in real compensation per measured R&D worker, the constructed R&D-worker headcount, and real GDP. Panel B applies an exact three-factor Shapley decomposition to $100 \times (s^I_{Lt}/s^I_{L0} - 1)$, using P_t/P_0, N_t/N_0, and Y_0/Y_t. The method averages marginal effects over all six factor orderings, so the three percentage-point contributions sum to the income share's percent change. P_t is not a fixed-quality wage index, and the GDP contribution is denominator scaling rather than a vintage allocation. Contributions are accounting allocations, not causal effects; worker data end in 2021.

Panel A shows that real compensation per measured worker rises by 116.0 percent and R&D-worker headcount rises by 166.9 percent through 2021. Real GDP rises by 189.4 percent over the same period. Panel B assigns 132.56 percentage points to real compensation per measured worker, 166.79 percentage points to worker headcount, and −200.23 percentage points to real GDP. These offsetting contributions yield a 99.12 percent increase in the income share.

The rising R&D workforce intensity of the economy, as shown in Figure 1 Panel B, is consistent with this decomposition. That share is N_t/E_t, where E_t is total employment. Growth in N_t raises both the workforce share and the R&D labor income share, holding their other components fixed. Real output enters only the latter calculation and offsets much of the combined growth in compensation per worker and headcount.

With our baseline assumptions, the negative GDP contribution is mechanical. Since $V_t = \kappa \times$

$D_t \times Y_t$, a larger output base raises the estimated value of a given productivity improvement. Thus, the denominator of the (R&D) income share grows with the level of final-goods income.

The rise in both compensation per R&D worker and R&D employment is consistent with an outward shift in demand for specialized workers when short-run labor supply is inelastic (e.g., Goolsbee 1998). Evolution in the R&D workers' bargaining power could move compensation in the same direction. Furthermore, Changes in worker quality, non-wage compensation, capital–labor complementarity, or measurement error would also move compensation. The decomposition above cannot discern between these explanations.

4. R&D Labor's Share of Costs

[INCOMPLETE]

5. Discussion

We find that, in the US over the past forty years, R&D labor has captured about 12% of the income it helps generate, and that share has roughly doubled from 8% in the 1980s to 16% in recent years. Alternative valuation assumptions change the level and timing, but every specification shows significant growth.

Why did the income share rise? For now we can only answer that question in the accounting sense. A decomposition assigns 132.56 percentage points to real compensation per measured worker and 166.79 percentage points to worker headcount. Growth in real GDP contributes −200.23 percentage points in the opposite direction.

The joint increase in compensation per worker and headcount admits several explanations. An outward shift in demand for specialized research labor, combined with an inelastic supply, can raise both compensation and employment (Goolsbee 1998). Changes in the way workers share the rents created by innovation (e.g., Van Reenen 1996; Kline et al. 2019) may also be evolving. The data do not settle between these, or any other, mechanisms.

References

Acemoglu, Daron, and Pascual Restrepo. 2018. "The Race between Man and Machine: Implications of Technology for Growth, Factor Shares, and Employment." *American Economic Review* 108 (6): 1488–1542.

Adams, James D. 1990. "Fundamental stocks of knowledge and productivity growth." *Journal of Political Economy* 98 (4): 673–702.

Aghion, Philippe, and Peter Howitt. 1992. "A Model of Growth Through Creative Destruction." *Econometrica* 60 (2): 323–351.

Aghion, Philippe, Benjamin F Jones, and Charles I Jones. 2019. "Artificial intelligence and economic growth." In *The economics of artifical intelligence: An agenda*, edited by Ajay Agrawal, Joshua Gans, and Avi Goldfarb: National Bureau of Economic Research Cambridge, Mass., USA.

Akcigit, Ufuk, Douglas Hanley, and Nicolas Serrano-Velarde. 2021. "Back to basics: Basic research spillovers, innovation policy, and growth." *The Review of Economic Studies* 88 (1): 1–43.

Arora, Ashish, Sharon Belenzon, and Lia Sheer. 2021. "Knowledge spillovers and corporate investment in scientific research." *American Economic Review* 111 (3): 871–98.

Arque-Castells, Pere, and Daniel F Spulber. 2022. "Measuring the private and social returns to R&D: Unintended spillovers versus technology markets." *Journal of Political Economy* 130 (7): 1860–1918.

Arrow, Kenneth. 1962. "Economic Welfare and the Allocation of Resources for Invention." In *The Rate and Direction of Inventive Activity: Economic and Social Factors*, 609–626: Princeton University Press.

Autor, David, David Dorn, Lawrence F. Katz, Christina Patterson, and John Van Reenen. 2020. "The Fall of the Labor Share and the Rise of Superstar Firms." *The Quarterly Journal of Economics* 135 (2): 645–709.

Azoulay, Pierre, Joshua S Graff Zivin, Danielle Li, and Bhaven Sampat. 2019. "Public R&D investments and private-sector patenting: Evidence from NIH funding rules." *Review of Economic Studies* 86 (1): 117–152.

Babina, Tania, Alex Xi He, Sabrina T Howell, Elisabeth Ruth Perlman, and Joseph Staudt. 2023. "Cutting the Innovation Engine: How Federal Funding Shocks Affect University Patenting, Entrepreneurship, and Publications." *The Quarterly Journal of Economics* 138 (2): 895–954.

Bloom, Nicholas, Charles I Jones, John Van Reenen, and Michael Webb. 2020. "Are ideas getting harder to find?" *American Economic Review* 110 (4): 1104–1144.

Bloom, Nicholas, Mark Schankerman, and John Van Reenen. 2013. "Identifying technology spillovers and product market rivalry." *Econometrica* 81 (4): 1347–1393.

Fieldhouse, Andrew, and Karel Mertens. 2023. "The returns to government R&D: Evidence from US appropriations shocks." Mimeo.

Fleming, Lee, Hillary Greene, G Li, Matt Marx, and Dennis Yao. 2019. "Government-funded research increasingly fuels innovation." *Science* 364 (6446): 1139–1141.

Goolsbee, Austan. 1998. "Does Government R&D Policy Mainly Benefit Scientists and Engineers?" *American Economic Review* 88 (2): 298–302.

Grossman, Gene M, and Elhanan Helpman. 1991. "Quality ladders in the theory of growth." *The Review of Economic Studies* 58 (1): 43–61.

Jaffe, Adam B. 1989. "Real effects of academic research." *American Economic Review* 79 (5): 957–970.

Jones, Benjamin F. 2009. “The burden of knowledge and the “Death of the renaissance man”: Is innovation getting harder?” *Review of Economic Studies* 76 (1): 283–317.

Jones, Benjamin F. 2025. “Artificial Intelligence in Research and Development.” Mimeo.

Jones, Benjamin F, and Larry Summers. 2022. “A Calculation of the Social Returns to Innovation.” In *Innovation and Public Policy*, edited by Austan Gooslbee and Benjamin F Jones.

Jones, Charles I. 1995. “R&D-based models of economic growth.” *Journal of Political Economy* 103 (4): 759–784.

Jones, Charles I, and John C Williams. 1998. “Measuring the social return to R&D.” *The Quarterly Journal of Economics* 113 (4): 1119–1135.

Kantor, Shawn, and Alexander Whalley. 2019. “Research proximity and productivity: long-term evidence from agriculture.” *Journal of Political Economy* 127 (2): 819–854.

Karabarbounis, Loukas. 2024. “Perspectives on the labor share.” *Journal of Economic Perspectives* 38 (2): 107–136.

Karabarbounis, Loukas, and Brent Neiman. 2014. “The Global Decline of the Labor Share.” *The Quarterly Journal of Economics* 129 (1): 61–103.

Kline, Patrick, Neviana Petkova, Heidi Williams, and Owen Zidar. 2019. “Who profits from patents? rent-sharing at innovative firms.” *The Quarterly Journal of Economics* 134 (3): 1343–1404.

Myers, Kyle R, and Lauren Lanahan. 2022. “Estimating spillovers from publicly funded R&D: Evidence from the US Department of Energy.” *American Economic Review* 112 (7): 2393–2423.

Nelson, Richard. 1959. “The simple economics of basic scientific research.” *Journal of Political Economy* 67 (3): 297–306.

Ricardo, David. 1817. *On the Principles of Political Economy and Taxation*.: J. Murray.

Romer, Paul M. 1990. “Endogenous technological change.” *Journal of Political Economy* 98 (5, Part 2): S71–S102.

Ryoo, Jaewoo, and Sherwin Rosen. 2004. “The Engineering Labor Market.” *Journal of Political Economy* 112 (S1): S110–S140.

Stern, Scott. 2004. “Do scientists pay to be scientists?” *Management Science* 50 (6): 835–853.

Van Reenen, John. 1996. “The Creation and Capture of Rents: Wages and Innovation in a Panel of U.K. Companies.” *The Quarterly Journal of Economics* 111 (1): 195–226.

Supplemental Appendix

Appendix A. Details on Data Construction

This appendix describes how the analysis files are constructed from the underlying surveys and national accounts. After the interpolation described below, the aggregate analysis file has one observation per year from 1981 through 2022. The business-industry source panel contains 34 survey years: odd years from 1981 through 1995 and every year from 1997 through 2022. All R&D quantities are assigned to the sector that performs the research, regardless of who funds it.

A.1. Sources and Measurement Conventions

Table A1 summarizes the inputs used in the paper. The federal R&D surveys measure domestic intramural R&D: work performed within the reporting organization in the United States. They exclude R&D performed abroad and, to avoid double counting, work purchased from another domestic performer when it appears in that performer's intramural expenditures.

TABLE A1
Principal Data Sources

Source	Role in the analysis
NCSES–Census SIRD, BRDIS, BRDS, and BERD surveys	Business R&D expenditures, compensation components, industries, and personnel
NSF HERD survey	Higher-education R&D expenditures and personnel
IPEDS	Higher-education research wages and fringe benefits
BEA R&D Satellite Account and NSF National Patterns	Government and nonprofit compensation and personnel; performer-specific R&D expenditures
Published NSF national total across all performers	Gapless series of total US R&D expenditures
OECD Main Science and Technology Indicators	Researcher full-time equivalents and total US employment
BEA National Income and Product Accounts	Nominal GDP, real GDP, and the GDP price deflator
Penn World Table 11	US total factor productivity

The analysis distinguishes nominal accounting ratios from real levels. Labor compensation and source-reported R&D expenditures are both nominal flows, so their ratio requires no deflator. For every dollar level, we use the BEA GDP price deflator and rebase the resulting series to 2022 dollars. The valuation model first produces real R&D-based income, $\kappa \times Y_t$.

Multiplying by the GDP price deflator gives the nominal denominator, $\kappa \times D_t \times Y_t$, which we compare with nominal compensation. Equivalently, both numerator and denominator can be expressed in 2022 dollars with the same deflator.

The source expenditure concepts differ across performing sectors and are accounting measures rather than harmonized economic costs. For example, business R&D expense includes depreciation and equipment charged to current expense, but excludes separately reported capital expenditures for long-lived facilities and equipment. It does not replace observed depreciation with the user cost of the assets employed in research. The accounting ratios in Section 2 therefore do not measure labor's economic share of R&D expenditures. Constructing that share requires a separate treatment of capital costs.

A.2. Business R&D

Survey vintages.. We join successive surveys conducted by NCSES and the Census Bureau: the Survey of Industrial Research and Development (SIRD), also published as *Research and Development in Industry*, through 2007; the Business R&D and Innovation Survey (BRDIS) for 2008–2016; the Business R&D Survey (BRDS) for 2017–2018; and the Business Enterprise Research and Development Survey (BERD) thereafter. We also use a hand-collected file for 1981–1989. The business total is observed in every survey year. In 2008–2010, firms report domestic intramural R&D costs in two pieces: costs funded by the company and costs funded by others. We add them to obtain the same total used in other survey years. The names and detail of the underlying cost items vary across survey vintages.

Compensation.. Business R&D labor compensation includes salaries and wages, employer fringe benefits, stock-based compensation, and temporary staffing. Salaries and wages are available throughout the sample. Fringe benefits are reported separately from 2001 through 2008 and combined with salaries and wages thereafter. Stock-based compensation and temporary staffing first appear in 2008. Temporary-staffing expenditures may include agency margins, so not all of this amount necessarily reaches workers.

We keep every component reported in a survey. If a component is absent, we impute it as a share of total intramural R&D expenditures. This component imputation is separate from the calendar-year interpolation of aggregate totals described below. For each missing component, we fit a logit-linear time trend to its observed expenditure share at the most detailed level available: the business total, major industry, detailed industry, or firm-size group. Multiplying the fitted share by reported expenditures gives the dollar amount. The baseline projects the estimated trend backward. A conservative sensitivity check instead holds the earliest observed share fixed.

The imputations make the compensation definition comparable across survey vintages. They do not provide direct evidence on fringe benefits, stock-based compensation, or temporary staffing before those items were reported. For these components, the reported years carry the most information.

A.3. Business Industries

We organize the industry data in a fixed hierarchy. Business R&D is first divided between manufacturing and nonmanufacturing, then among 26 mutually exclusive industries. The manufacturing branch contains 18 groups, generally at the three-digit NAICS level; the nonmanufacturing branch contains eight, generally at the two-digit level. Both branches include an "Other" residual. The industries add to their branch totals, and the two branches add to the published business total.

The shift from SIC to NAICS in 1999 is the main break in the industry data. We retain a pre-1999 series when it is comparable with its NAICS-era counterpart. Four electronics and instruments groups are not comparable, so we reproject their pre-1999 observations. Where no consistent series exists, we fit a logit-linear trend to the industry's NAICS-era share and project it backward. We assign the projection only to the part of the observed manufacturing or nonmanufacturing total not already accounted for by reported industries. Three combined SIC categories remain combined because their NAICS children are not observed. Construction, retail trade, and electronic shopping are too sparsely reported to support stable panels and enter the appropriate "Other" residual.

The business total is observed in 1981 and 1983, but a complete split between manufacturing and nonmanufacturing begins in 1985. Nonmanufacturing expenditures are observed in 1983 and projected for 1981; manufacturing is the residual. We divide compensation between the two branches so that they sum to published business compensation. At the industry level, observed labor ratios outside 0.15–0.92 are replaced by the sector-specific logit-share prediction. When compensation is missing, we calculate it from the projected labor share and bound it by the industry's expenditures. We then rescale industry compensation proportionally to the branch total, repeating the calculation only when a cell reaches a bound in $0 \le W_{jt} \le R_{jt}$. This preserves aggregate compensation without allowing compensation to exceed expenditures. A flag records whether an industry's spending is reported or constructed.

The resulting panel adds up, but a projected cell is still a projection. This matters especially for nonmanufacturing before 1999: information and professional services are entirely projected in those years. None of these choices affects the aggregate business results because they leave the observed business total unchanged. For that reason, the industry

figure focuses on broad groups and later movements rather than narrow comparisons before 1999. We select its four named business industries once, based on cumulative real R&D expenditures over the full sample, and hold that membership fixed over time.

A.4. Higher Education

HERD reports total intramural R&D expenditures at colleges and universities. IPEDS provides institution-level research wages and, for many institution-years, research fringe benefits. When research fringe benefits are missing but the institution reports total wages and fringe benefits, we apply its institution-wide fringe-to-wage ratio to research wages. We then add wages and fringe benefits across institutions. A year is not treated as a direct observation of the national labor ratio if institutions with unresolved wage or fringe omissions account for more than 5 percent of national research expenditures.

From 1987 onward, the preferred series uses the direct IPEDS wage-plus-fringe estimate whenever it passes this screen. For 1981–1986, we backcast the higher-education labor ratio with a state-space model. The latent ratio follows a random walk and is measured with error using the direct wage-and-fringe ratio and a longer wages-only proxy. We do not use this model to fill later gaps. The direct series remains missing in 1997, 2000, 2001, and 2015; those source observations enter only the later aggregate interpolation and residual completion. HERD begins reporting salaries, wages, and fringe benefits directly in 2010. We retain the HERD values as a check but use IPEDS in the preferred series because it provides a consistent measure farther back in time.

A.5. Government and Nonprofit R&D

NSF National Patterns reports R&D expenditures by performer. Government combines federal and nonfederal government; nonprofit performers remain separate. The BEA R&D Satellite Account provides compensation and employment for 2012–2023. We remove colleges and universities from BEA's government category so that they are not counted again in higher education.

For each of the government and nonprofit sectors, we calculate the expenditure-weighted average ratio of BEA compensation to NSF expenditures over 2012–2023. Before 2012, compensation equals that sector-specific average multiplied by its observed R&D expenditures. The early series therefore captures changes in sector size but, by construction, contains no within-sector movement in the labor ratio. Business temporary-staffing costs do not have a direct counterpart in BEA employee compensation, another reason to avoid interpreting small cross-sector differences as exact.

A.6. R&D Workers and the Workforce Denominator

The personnel sources use different concepts. OECD reports researchers in full-time-equivalent units; the business surveys report scientists and engineers in early years and a broader R&D-worker headcount in later years; HERD and the R&D Satellite Accounts report sector headcounts. We construct one all-sector headcount in two stages.

First, we build a continuous series of all-sector researcher full-time equivalents. OECD reports the all-sector total from 1991 onward. For the 1980s, we combine available higher-education and government researcher series and scale them using the average residual share outside those two sectors over the 1991–1999 overlap. This residual principally captures business researchers but also includes nonprofit researchers. Interior gaps are linearly interpolated. Second, we convert researcher full-time equivalents to the broader R&D-worker headcount. In 2012–2021, when business, higher-education, government, and nonprofit headcounts are jointly available, we sum the four sectors and calculate their average ratio to OECD researcher full-time equivalents. We apply that conversion ratio to the 1981–2011 researcher series. The resulting measure runs through 2021.

The denominator of the workforce share is OECD total US employment in persons. The numerator is a constructed headcount, not a count of identical workers or a quality-adjusted labor input. Moreover, the compensation and headcount series do not measure perfectly matched worker populations: temporary-staffing expenditures enter compensation, while the constructed headcount need not include the same temporary workers. Compensation per measured worker is therefore neither an average wage nor a fixed-quality price index. Changes in occupation, education, hours, sector mix, and worker quality may also affect it.

A.7. National Expenditures, Output, and Productivity

Figure 1 uses NSF's published national R&D total, which provides a gapless series, divided by nominal GDP from the BEA. All compensation-based figures use the bottom-up sum of business, higher-education, government, and nonprofit series. The income calculation combines this compensation total with nominal and real GDP and the GDP price deflator from the BEA and the US TFP index from Penn World Table 11.

The bottom-up R&D expenditure total is missing in the even years of the early biennial business surveys. The bottom-up compensation total has those gaps and four additional years in which the higher-education wage-quality screen removes the direct labor split. We linearly interpolate these aggregate nominal totals and retain flags for every filled observation; the compensation series contains 12 flagged years. Real totals are computed only after this interpolation. In the sector lines of the income figure, whichever broad

component is unavailable in a flagged year is the residual between the interpolated total and the observed counterpart. No such residual completion is used in the sector-level accounting-cost figures.

A.8. Reconciliation and Principal Limitations

The construction code applies a hierarchy of accounting checks. Within business R&D, industry expenditures and compensation close to their manufacturing or nonmanufacturing parent, and the two parents close to the business total. Across performing sectors, business, higher education, government, and nonprofit totals close to the aggregate. The checks also require compensation not to exceed expenditures in any industry cell and prevent retired or unknown industry labels from reappearing. We retain and flag the small discrepancies in a few published nonmanufacturing subtotals rather than alter the source data to make them disappear.

Several features of the construction matter for interpretation. Early business compensation relies on backcasts of nonwage components. Before 1999, industry composition is partly projected, especially within nonmanufacturing. Higher-education compensation is backcast before 1987 and missing in four later source years; government and nonprofit labor ratios are held fixed before 2012. The aggregate compensation series interpolates flagged years and sometimes recovers one side of the sector split as a residual. The worker series uses overlap-period ratios to convert researcher full-time equivalents into a broader headcount. Finally, compensation and headcount do not cover exactly the same workers. These choices allow us to compare broad movements over a long period, but not to put the same weight on individual annual changes or narrowly defined industries early in the sample.