Federal Reserve Bank of Dallas

# The Returns to Government R&D: Evidence from U.S. Appropriations Shocks

Andrew J. Fieldhouse and Karel Mertens

**Working Paper 2305** **May 2023 (Updated November 2024)**

Research Department

https://doi.org/10.24149/wp2305r2

# The Returns to Government R&D: Evidence from U.S. Appropriations Shocks[*]

Andrew J. Fieldhouse[†] and Karel Mertens[‡]

First version: May 5, 2023
This version: November 17, 2024

**Abstract**

Based on a narrative classification of all significant postwar changes in R&D appropriations for five major federal agencies, we find that an increase in nondefense R&D appropriations leads to increases in various measures of innovative activity and higher business-sector productivity in the long run. We structurally estimate the production function elasticity of nondefense government R&D capital using the SP-IV methodology of Lewis and Mertens (2023) and obtain implied returns of 140 to 210 percent over the postwar period. The estimates indicate that government-funded R&D accounts for one-fifth of business-sector TFP growth since WWII, and imply substantial underfunding of nondefense R&D.

**JEL Classification:** E62, O38, O47.

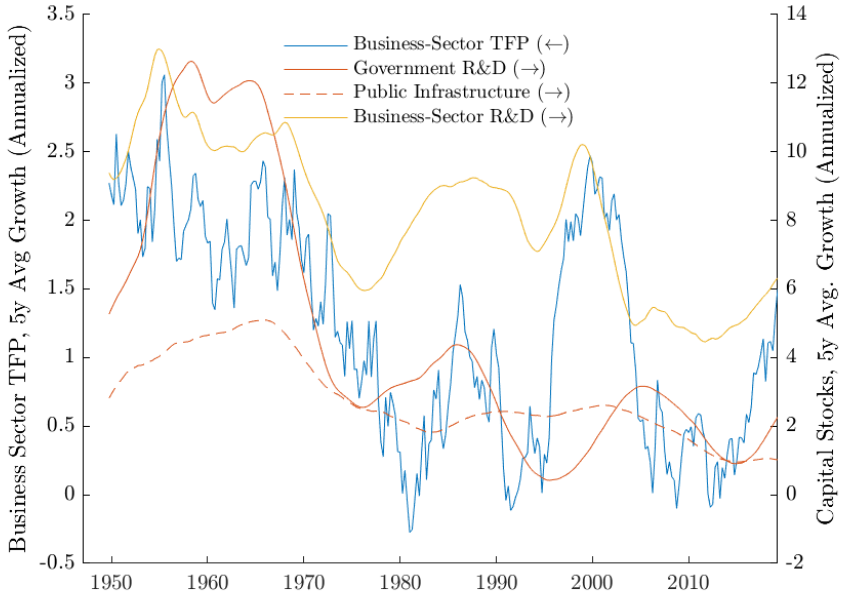
**Keywords:** Government R&D, Productivity, Growth, Narrative Analysis.

[*] We are grateful to our discussants, Pete Klenow and Heidi Williams, as well as to Peter Beall, Diego Comin, Alexander Field, Soroush Ghazi, Joseba Martinez, Paolo Surico, Ana Maria Santacreu, and Johannes Wieland for helpful comments. We also thank seminar participants at Arizona State University, Alabama University, Barcelona Summer Forum, Central Bank of Ireland, Clark University, Congressional Budget Office, Cornell Economics Alumni workshop, Federal Reserve Banks of Atlanta and Dallas, Ghent University, Leuven University, Liberal Arts Macroeconomics Conference, London Business School, Lundquist College of Business, Mays Business School, Middlebury College, Midwest Macro Conference, NBER SI Innovation, New Thinking in Industrial Policy Conference, Notre Dame, Northwestern University, Oxford University, Texas A&M University, University of Houston, UT Austin, and the Yale Cowles Conference on Macroeconomics for helpful feedback. Roni Lezama, Annie Midthun, Tai Nguyen, Michaela Sullivan, and Nicholas Wagg provided excellent research assistance. Andrew Fieldhouse gratefully acknowledges funding received from Middlebury College. The views in this paper are those of the authors and do not necessarily reflect the views of the Federal Reserve Bank of Dallas or the Federal Reserve System.

[†] Andrew J. Fieldhouse, Mays Business School, Texas A&M University.

[‡] Contact: Karel Mertens, Federal Reserve Bank of Dallas and CEPR, mertens.karel@gmail.com, tel: +(214) 922-6000.

Figure 1: Growth in Business-Sector TFP, R&D Capital, and Public Infrastructure




*Notes*: Centered five-year moving average annualized growth rates based on quarterly data. Business-sector TFP is the utilization-adjusted measure of Fernald (2012). Public infrastructure consists of nondefense structures and equipment. The definition of R&D capital includes a capitalization of expenditures for software development. See Appendix A for variable definitions. *Sources*: BEA, Fernald (2012).

With the exception of a brief period in the late 1990s and early 2000s, aggregate U.S. productivity growth has slowed markedly since the late 1960s. Figure 1 shows that this slowdown coincides with a decline in public investments in research and development (R&D).[1] The causality underlying this relationship, however, is far from clear, and as Figure 1 shows, higher growth in business R&D capital or public infrastructure prior to the 1970s are plausible alternative contributing factors.

Several significant empirical challenges need to be overcome in order to isolate the causal role of government R&D in driving innovation and productivity growth. Any productivity spillovers likely occur only after long and uncertain lags. Various potential channels for reverse causality need to be accounted for, since policymakers' decisions to boost or cut R&D funding could be influenced by a wide range of factors with independent effects on innovation. Aggregate estimates must also be interpreted with care, as more government funding can impact private spending on R&D or other productivity-enhancing public investments.

In this paper, we propose a novel empirical strategy to estimate the aggregate dynamic effects of changes in government R&D spending and to identify direct versus indirect productivity effects. Because the lags between spending decisions and actual outlays are often long, the starting point of our analysis is a new dataset of all postwar appropriations enacted for the budgetary accounts funding R&D at the major federal agencies: The Department of

[1]See also Mazzucato (2018), Gruber and Johnson (2019), and Bloom et al. (2019), among others.

Defense (DOD), Department of Energy (DOE), National Aeronautics and Space Administration (NASA), National Institutes of Health (NIH) within the Department of Health and Human Services, National Science Foundation (NSF), and their historical precursors. To guard against reverse causality, we perform a narrative classification of all major changes in federal R&D appropriations for these agencies to construct measures that, after conditioning on a suitable set of controls, are largely unanticipated and plausibly free of confounding influences. We use the narrative measures in long-horizon Jordà (2005) local projections with quarterly postwar data to estimate the dynamic causal effects of shocks to R&D appropriations on aggregate TFP and various other indicators of innovative activity.

The knowledge spillovers from defense and nondefense R&D are likely quite different, if only because military R&D work is often classified to maintain military superiority, limiting spillovers. For this reason, we distinguish throughout the analysis between defense and nondefense R&D. We find that a positive shock to nondefense R&D appropriations robustly leads to a delayed and gradual increase in business-sector TFP that becomes highly statistically significant at long forecast horizons (8 to 15 years). For a shock that induces a one percent increase in government R&D capital, our baseline estimates show eventual TFP increases of about 0.2 percent. Positive shocks to nondefense R&D also induce increases in various indicators of innovative activity, such as patent grants, the number of doctoral recipients in STEM fields, the number of researchers engaged in R&D, or the number of technology publications. In contrast, we find little evidence that a positive shock to defense R&D leads to any persistent productivity increases, at least not within horizons of 15 years.

To better understand the estimated TFP responses, we investigate various decompositions of the spending changes that occur following shocks to R&D appropriations. As emphasized by Akcigit et al. (2020), public investments that focus more heavily on producing basic knowledge can create important complementarities with private research investments and have large spillovers. We find that nondefense shocks lead to relatively larger increases in funding for basic research, and to particularly persistent increases in funding for research performed within government agencies and at universities. The majority of the increase in nondefense R&D funding, in terms of dollars, stems from higher appropriations for NASA, followed by the NIH. Defense shocks instead mostly result in increased funding for development and product improvement, with more of the work performed by businesses. Some of these compositional differences could help explain why we do not detect meaningful spillovers from defense R&D. However, it is also possible that the spillovers from defense R&D simply take longer to materialize than those from nondefense R&D and, therefore, it would require longer samples to detect them reliably on the basis of our methodology.

The bulk of the increase in publicly funded R&D work following either shock is performed outside of the federal agencies, i.e., at private businesses, public-private R&D centers, or universities. However, we find that positive shocks to both defense and nondefense R&D appropriations also lead to higher private investment expenditures in R&D. As in the theoretical framework of Akcigit et al. (2020), this suggests that private and public R&D investments act as complements rather than substitutes. However, the increases in privately funded R&D following a nondefense shock are relatively small, and larger transitory increases following a defense shock are reversed in the longer run. We find that one channel through which a positive nondefense shock likely has important additional indirect effects on productivity is a gradual expansion of public infrastructure funded by state and local governments. This expansion is broad-based, with the largest increases in education structures (schools and universities), followed by roads and power, water, and sewer systems.

In order to isolate the direct productivity effects of government R&D, we formulate an aggregate production function with public infrastructure and government R&D capital as separate arguments, and we structurally estimate the elasticity to government R&D capital. Our identification strategy relies on two key steps. First, we use available estimates of the production function elasticity of public infrastructure to remove its contribution to business-sector TFP growth. In the second step, we use the SP-IV estimator of Lewis and Mertens (2023) to estimate the production function elasticities of defense and nondefense government R&D capital. Intuitively, this estimator is a GMM estimator that obtains the elasticity as the value that best fits the relationship between the estimated responses of government R&D capital and (infrastructure-adjusted) TFP to the R&D appropriations shocks. Based on the response to a nondefense shock, the point estimates of the production function elasticity to government R&D capital across various specifications lie within a relatively tight range around a value of 0.11, and these estimates are generally highly statistically significant under weak-instrument-robust inference procedures.[2] In contrast, the results for defense R&D are inconclusive, as the estimates vary greatly across specifications and are very imprecise.

In a growth accounting exercise, our estimates imply that nondefense government R&D accounts, on average, for at least one-fifth of business-sector TFP growth in the postwar period. Despite the fact that the government invests significantly less in R&D than in infrastructure, the contribution of government R&D to TFP growth is consistently of a similar magnitude to, and frequently greater than, the contribution of public infrastructure. Depending on the assumed value of the public infrastructure elasticity, slower growth in all forms of public capital explains 0.36 to 0.43 percentage points of the TFP slowdown of around one percentage point after the 1960s. Our findings indicate that the slower growth in government R&D was equally important, if not more so, than the slowdown in public infrastructure investment for the deceleration in U.S. TFP growth.

Finally, we calculate the rate of return to nondefense government R&D, both indirectly from the elasticity estimates and directly from SP-IV estimates in regressions of TFP growth on the ratio of net R&D investment to output. Depending on the method of calculation and specification, we obtain rates of return on nondefense R&D in a range of 140 to 210 percent

[2]The value of 0.11 is for total government R&D capital, and translates to an elasticity to nondefense R&D capital of 0.06, given that nondefense R&D averages about one-half of total government R&D in the postwar sample.

under a Cobb-Douglas assumption. These estimates are considerably higher than similar ones for the return on public infrastructure. Our findings, therefore, point to a misallocation of public capital and substantial underinvestment in nondefense R&D.

This paper contributes to a large empirical literature estimating 'social' returns to R&D, i.e., returns that include various spillovers on other firms or industries, which are typically found to well exceed the normal return on other investments.[3] Firm or industry-level studies, however, are restricted in the scope of spillovers and general equilibrium effects that can be captured. While aggregate data are better suited for estimating the concept of a 'social' return, the main challenge is causal identification. Our paper proposes a strategy for causal identification with aggregate data in the context of government-funded R&D.

A number of existing studies focus on private spillovers of specific government R&D programs. For instance, Azoulay et al. (2019) find that NIH spending causes the development of more private patents; Myers and Lanahan (2022) find large private R&D spillovers from the DOE's Small Business Innovation Research program; Kantor and Whalley (2024) find persistent manufacturing value added spillovers from local NASA R&D spending during the moon mission; Gross and Sampat (2023) document how R&D programs during WWII fueled the postwar growth of technology clusters and spurred innovation in the long term; Moretti et al. (2021) find positive spillovers from defense R&D on private R&D and productivity growth in a panel study of OECD countries; and Babina et al. (2023) find that negative shocks to federal grants for research at universities reduce high-tech entrepreneurship and publications. Each of these studies provides evidence for some of the spillovers that we aim to measure collectively.

A closely related recent paper by Dyèvre (2024) uses firm-level Compustat data matched to patents to study the aggregate spillovers from federally funded R&D to firms' innovative work and output, exploiting shift-share and patent examiner leniency instrumental variables. Despite the very different identification approach, Dyèvre (2024) finds large, positive causal effects on firms' productivity and innovative efforts. Based on a structural model calibrated to the empirical findings, Dyèvre (2024) concludes that the decline in government R&D spending can account for about one-third of the postwar slowdown in U.S. TFP growth.

Our paper is also related to several recent studies of the longer-run macroeconomic effects of fiscal policy shocks. Cloyne et al. (2024), for example, find that a corporate tax cut leads to increases in R&D spending by businesses, as well as longer-run increases in TFP. Antolin-Diaz and Surico (2024) study the long-run effects of military spending shocks and find that these shocks lead to long-run increases in output and productivity. Consistent with our results, the authors argue that the long-run effects arise because of increases in the share of government spending going to R&D. The authors additionally explore the effects of a distinct public R&D shock identified by maximizing the forecast error variance of government R&D spending over horizons of up to one year, and like us, they find that a positive shock causes a

[3] See Hall et al. (2010) or Jones and Summers (2022) for overviews of the evidence.

gradual increase in TFP. De Lipsis et al. (2022) also study the effects of shocks to government R&D spending, in their case identified with short-run restrictions similar to Blanchard and Perotti (2002). As we do, they find that government R&D crowds in private investment and raises output in the long run. Different from Antolin-Diaz and Surico (2024) or De Lipsis et al. (2022), we focus on shocks to R&D appropriations rather than R&D spending, use a narrative identification scheme, and distinguish between defense and nondefense government R&D. Despite the methodological differences, it is reassuring that our conclusions regarding the potential for government R&D spending to boost economic growth are broadly similar.

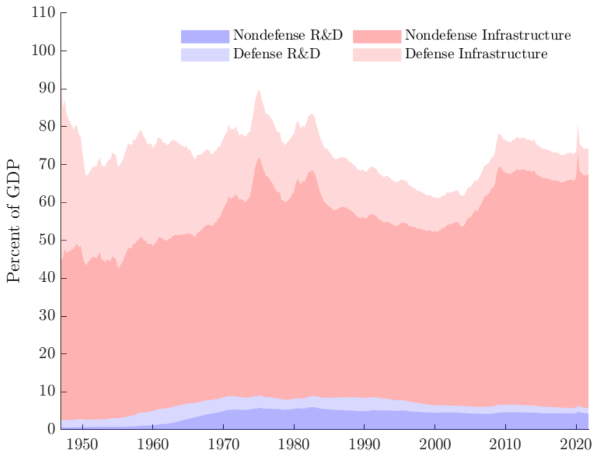
Finally, this paper contributes to the literature on the productivity effects of public capital, see e.g., Bom and Ligthart (2014) and Ramey (2021) for surveys. Since the early contributions of Aschauer (1989) and Munnell (1990), this literature has mostly focused on (nondefense) public infrastructure. Our paper presents estimates of the production function elasticity of government R&D capital that can be used to separately study the role of intangible public capital in quantitative growth models. Our empirical findings can also inform budgetary analyses of federal investments (e.g., CBO 2016; CBO 2021) by addressing some key outstanding challenges in evaluating the impact of federal R&D investments; see Campbell and Shirley (2018), Elmendorf et al. (2024), and Gullo et al. (2024).

## I. Measurement, Definitions, and Facts

The measures of public capital are based on data from the Bureau of Economic Analysis (BEA).[4] We distinguish between (i) defense non-R&D capital (defense-related equipment and structures), (ii) public infrastructure (federal nondefense and state and local government equipment and structures), (iii) defense R&D capital, and (iv) nondefense R&D capital (federal, state, and local government). Our definition of R&D capital includes a capitalization of expenditures for software development and therefore corresponds to the concept of 'intellectual property' for the government sector in the National Income and Product Accounts (NIPAs); we use the term 'R&D capital' as such throughout the rest of the paper. We refer to the aggregate of (iii) and (iv) as 'government R&D capital'. NIPA R&D expenditures are measured by source of funding, so government R&D capital includes federally-funded 'contract R&D' performed by firms, universities, nonprofits, and public-private partnership 'R&D centers' (e.g., the Lawrence Livermore National Laboratory). Figure 2 plots the quarterly time series of public capital and its subcomponents as a ratio of GDP. As is clear from the figure, government R&D capital is relatively small compared to other types of public capital, with nondefense and defense R&D capital averaging 3.9 percent and 2.7 percent of GDP, respectively, over the postwar period. For comparison, business-sector R&D capital averages 6.3 percent of GDP, see Appendix B.

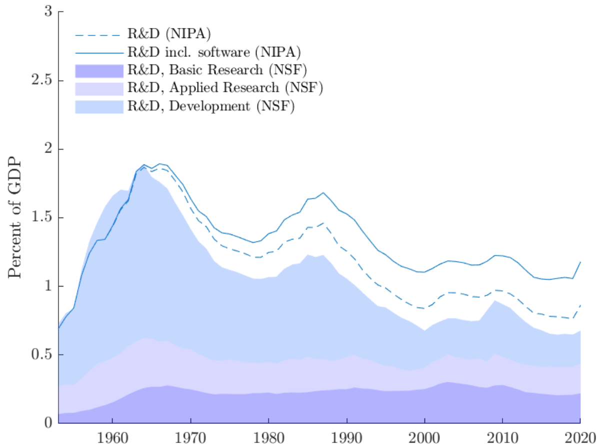
[4]We use NIPA gross investment data to construct quarterly series of government fixed asset values (at real cost) that are consistent with the BEA's annual Fixed Asset Account; see Appendix A for details. Appendix E.5 considers R&D capital measures based on different depreciation rates than those used by the BEA.

Figure 2: Composition of Public Capital




*Notes*: R&D capital includes software. Infrastructure consists of structures and equipment. See Appendix A for variable definitions. *Source*: BEA.

Figure 3: Government R&D Expenditures




*Notes*: Fiscal year NCSES data are converted to calendar years and exclude R&D plant. *Sources*: BEA; NCSES, National Patterns of R&D Resources (Tables 7, 8, and 9).

The expenditure data underlying the BEA measures of R&D capital are sourced from annual surveys conducted by the NSF's National Center for Science and Engineering Statistics (NCSES). Unlike the NIPA data, NCSES data on R&D spending are available by funding agency, performing sector, and type of research activity. The NCSES defines R&D as the "creative and systematic work undertaken in order to increase the stock of knowledge... and to devise new applications using available knowledge" (Moris and Pece 2022, p. 15). The NCSES separates R&D spending into three types: basic research, applied research, and experimental development work. Basic research is defined as "experimental or theoretical work undertaken primarily to acquire new knowledge of the underlying foundations of phenomena and observable facts" whereas experimental development work is defined as "creative and systematic work, drawing on knowledge gained from research and practical experience, which is directed at producing new products or processes or improving existing products or processes." In between, applied research is defined as "original investigation undertaken in order to acquire new knowledge... directed primarily towards a specific practical aim or objective" (Moris and Pece 2022, p. 17).

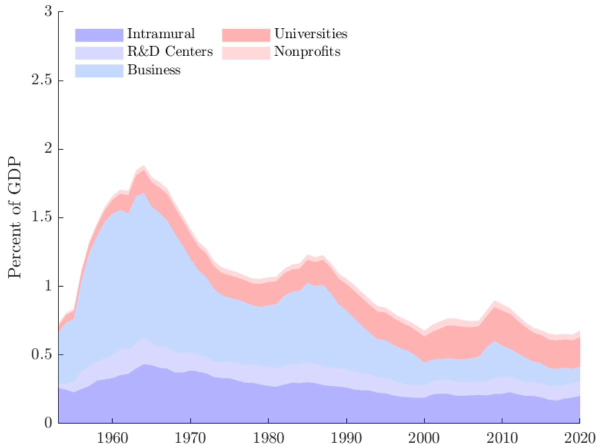
Figure 3 plots the NCSES measures of government R&D spending by type, along with NIPA totals for comparison. Spending on basic and applied research each average 0.23 percent of GDP over the sample, while experimental development averages 0.61 percent. Spending on basic research is considerably larger than that of the private sector, which instead spends relatively more on applied research and development.[5] As emphasized in Akcigit et al. (2020), this compositional difference suggests that distinguishing between private and public R&D spending is potentially important.

As Figure 3 shows, NCSES R&D expenditure data do not align perfectly with the cor-

[5]Private expenditures on basic research (development) average 0.14 (0.97) percent of GDP, see Appendix B.

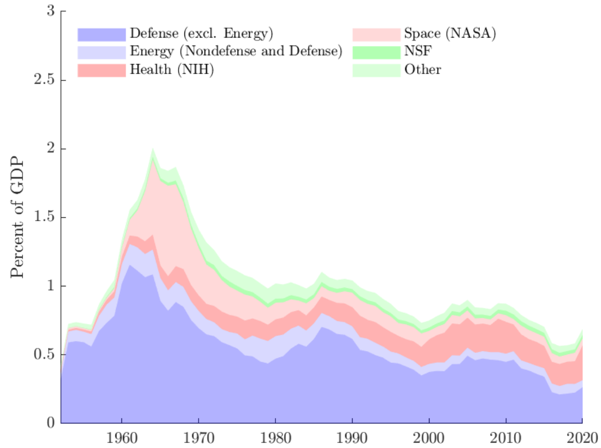
Figure 4: Government R&D Expenditures

(a) By Performing Sector




*Notes*: Fiscal year data are converted to calendar years. *Source*: NCSES, National Patterns of R&D Resources (Table 6).

(b) By Federal Agency




*Notes*: Federal R&D outlays by agency excluding R&D plant. Fiscal year data are converted to calendar years. *Source*: NCSES, Survey of Federal Funds for R&D (Various tables).

responding NIPA series, as the BEA adjusts the NCSES source data and uses additional budgetary data to match required NIPA concepts. 'Software development,' in particular, is a broader concept in the NIPAs and includes various non-experimental development expenditures.[6] Also, not all spending labeled research or development in other data sources necessarily flows exclusively into the NIPA measure of government R&D expenditures. For example, DOD spending on 'operational systems development' is mostly classified by the BEA as equipment. Similarly, 'R&D plant', i.e., spending on major research facilities and equipment, is also mostly recorded as investment in equipment or structures by the BEA.

Figure 4 plots NCSES data on government R&D spending broken out by performing sectors and the major funding agencies. Panel (a) shows that most government R&D spending funds activity performed by private businesses, universities, or public-private R&D centers, as opposed to 'intramural' R&D conducted within the federal agencies. At the height of the Cold War, most government-funded R&D was performed by businesses, but the share has declined since, and a steadily growing share is performed at universities.

On average, more than 90 percent of postwar public funding for R&D occurs at the federal level; the remainder consists mainly of state and local funding for research at universities. Panel (b) in Figure 4 provides a breakdown of federal R&D spending by agency. Early in the Cold War, DOD and NASA accounted for the bulk of federal R&D spending, and much of the decline in overall funding since the late 1960s results from Congress scaling back funding for these agencies after the moon landing and the deployment of the nuclear triad. Another major source of funding is DOE and its historical precursors, covering both defense activities (e.g., nuclear weapons and naval propulsion) and nondefense activities

[6] NIPA software development excludes software embedded in other products, e.g., computers or cars.

(e.g., civilian energy and physics research, much of which is conducted by the National Laboratories). In recent decades, NIH funding for basic and applied health research has gradually grown in importance. The final agency engaged in significant R&D funding is the NSF, which largely provides grants for basic research at universities. Private-sector spillovers from the R&D work funded by these five agencies have been extensively documented in qualitative work.[7] Other federal agencies also fund R&D work but in much smaller amounts.

## II. Measuring Exogenous Variation in Government R&D Spending

Our strategy for identifying the causal effects of government R&D spending is based on novel empirical measures of exogenous variation in federal funding for R&D. As is well known, an important identification concern is that changes in fiscal policy are often anticipated, and mistiming the arrival of fiscal news can be misleading (Ramey 2011; Mertens and Ravn 2013; Leeper et al. 2013; Brunet 2023). To address this concern, we rely on time series of all enacted appropriations for future federal R&D expenditures, and not just on current R&D expenditures as in Antolin-Diaz and Surico (2024) or De Lipsis et al. (2022).

The other immediate identification concern is that policy changes reflect systematic reactions by policymakers to macroeconomic developments that independently affect innovative activity and aggregate productivity. We take a two-step approach to isolating changes in appropriations that are plausibly uncorrelated with other influences on productivity and innovation. First, we adopt a narrative identification strategy and—on the basis of an extensive analysis of historical sources—retain only those changes in appropriations that are not motivated by short-run macroeconomic considerations. Second, to guard against the possibility that R&D policy responds systematically to other longer-term drivers of productivity, we embed the narrative measures in empirical models that remove predictable variation in future productivity through a wide variety of lagged controls at a quarterly frequency. Before we describe the methodology in full detail, the rest of this section first discusses the appropriations data and the narrative measures used for identification.

### A. Data on Appropriations for R&D

As the overwhelming majority of government R&D funding is at the federal level, we focus on congressional appropriations for R&D activities. To measure federal R&D appropriations, we rely on information in the *Budget of the U.S. Government* and its appendices. A companion paper, Fieldhouse and Mertens (2023), documents all enacted appropriations funding R&D activities at federal agencies for all fiscal years from 1947 to 2019. To keep the data collection manageable, we only consider the budget accounts for the five major

[7]See Mazzucato (2018) on the various technologies needed for an iPhone (e.g., GPS, internet, touchscreen, voice recognition) originating from government R&D funding or Gruber and Johnson (2019) on the Human Genome Project (funded by NIH and DOE) yielding spillovers across industries (e.g., agriculture, bio-fuels, pharma, and food processing).

federal agencies discussed in Figure 4: DOD, DOE, NASA, NIH, and NSF. Together, these five agencies typically account for around 90 percent of total federal R&D spending. For each agency, we obtain the appropriations from the 'Budget Authority' (BA) or—prior to the introduction of BA as a budgeting concept—the 'Appropriation (adjusted)' line item for each R&D account. The series we construct accounts for supplemental appropriations, subsequent transfers between accounts, or sequestration cuts. To match the defense and nondefense categories in the NIPAs, we split DOE appropriations into defense and nondefense functions using budget account-level data. We use primary sources to date the changes in appropriations to the quarter they take effect, either at the start of the fiscal year or when the appropriations bill was subsequently enacted. When multiple bills provide appropriations for one agency in the same fiscal year but take effect in different quarters, we use the first quarter in which a bill significantly changes real R&D funding.

### B. Narrative Classification

As mentioned earlier, one potential source of endogeneity is that R&D appropriations may be correlated with the business cycle. Comin and Gertler (2006) and Bianchi et al. (2019), for example, argue that expansionary business cycle shocks raise aggregate productivity at longer horizons through endogenous growth channels, while Ilzetzki (2024) shows that high capacity utilization during WWII spurred innovation out of necessity. Additional room in government budgets during booms may lead to increases in R&D appropriations without contributing to any longer-run productivity effects of that boom. On the other hand, appropriations for R&D may also rise in recessions if they are systematically folded into larger fiscal stimulus packages, e.g. as in the American Recovery and Reinvestment Act of 2009 (ARRA) or the Coronavirus Aid, Relief, and Economic Security Act of 2020.

To address the possible cyclical endogeneity of R&D appropriations, we could regress on lags of quarterly cyclical indicators and appeal to lags in the policymaking process for identification, as in Blanchard and Perotti (2002). However, this may not suffice to remove all cyclical influences, so we additionally use a narrative analysis to classify all policy changes as either 'exogenous' or 'endogenous' to short-run macroeconomic developments. The narrative analysis, detailed in Fieldhouse and Mertens (2023), covers each of the five agencies (separating the defense and nondefense functions of DOE) and all fiscal years from 1947 to 2019 with 'significant' changes in real appropriations, defined as year-over-year increases of at least 5 percent, or decreases of at least 2.5 percent. We focus on larger changes as legislative intent is easier to infer for more meaningful policy changes, and because this significantly reduces the number of agency-fiscal year pairs to analyze.

In total, we classify 257 significant policy changes, with roughly two-thirds involving increases in (real) appropriations. For each of these, we rely on various primary and secondary sources for context and motivation. Specifically, we study the *Budget of the United States*

*Government*, the *State of the Union Address*, and any related presidential signing statements, veto statements, or other speeches to learn the administration's budgetary priorities and specific goals for R&D policy. To infer legislative intent, we analyze the House and Senate Appropriations Committee reports that accompany each appropriations bill, as well as related committee hearings. We also cross-check every policy change with *CQ Almanac*'s analyses of the related appropriations bill(s) and any relevant authorizing legislation, as well as their overview of that year's budget cycle. Finally, we scan for newspaper coverage of the relevant appropriations bills, primarily *The Washington Post*, *The New York Times*, and *The Wall Street Journal*. Based on a close reading of these sources, we classify every significant policy change as either 'exogenous' or 'endogenous' to short-run cyclical concerns.

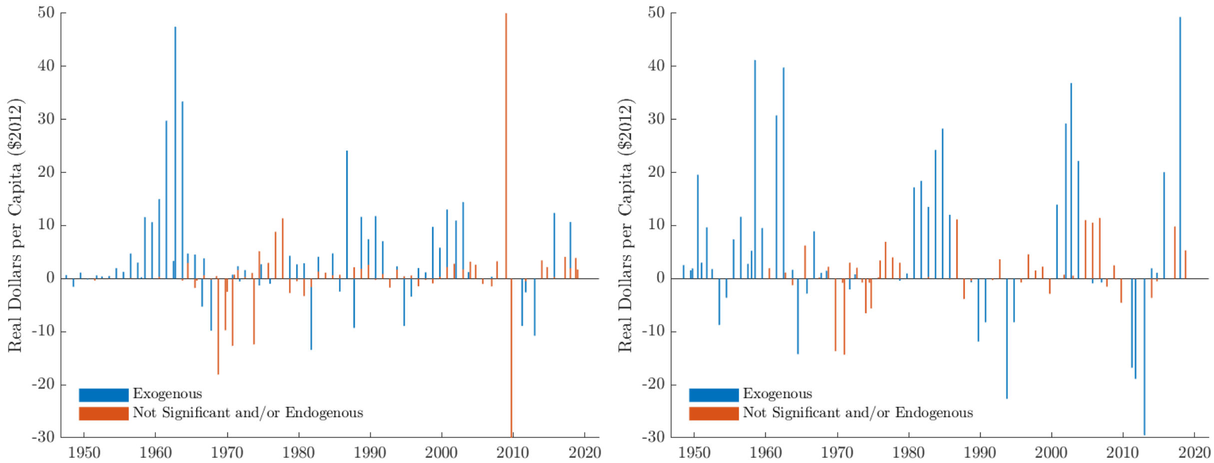
Based on the narrative evidence, we classify 46 of the 257 policy changes as endogenous. Several of these reflect increases in R&D spending that are part of broader fiscal stimulus packages (e.g., the increase in NIH funding under ARRA in 2009); some are motivated by energy shocks (e.g., funding for the Energy Research and Development Administration in 1974 in response to oil shocks); others are cuts that are part of broader austerity measures intended to curb short-run inflationary pressures (e.g., cuts to the R&D budgets of NASA, NIH, and the NSF early in the Reagan administration). We also classify any policy changes resulting from the expiration of temporary fiscal stimulus as endogenous (e.g., the reduction in NIH funding in 2010 from the expiration of ARRA funding).

For most of the significant changes to R&D appropriations, however, we find little evidence for short-run macroeconomic objectives as a significant motivating factor. Consistent with the qualitative accounts of Mazzucato (2018) and Gruber and Johnson (2019), the majority of the changes in federal R&D funding are mission-oriented and have specific longer-run strategic objectives, e.g., developing nuclear-powered submarines, developing new energy technologies that reduce carbon emissions, putting a man on the moon, reducing deaths from cancer or heart disease, or boosting the number of scientists and engineers graduating from American universities. The reasons for pursuing these objectives vary considerably. For defense R&D, examples include concerns about the adequacy of strategic capabilities relative to geopolitical rivals (e.g., the ballistic missile gap with the Soviet Union), policy preferences of a new administration (e.g., Reagan's military buildup), evolving national security threats (e.g., the Global War on Terror), or ratifying or exiting non-proliferation treaties (e.g., exiting the Anti-Ballistic Missile Treaty). For nondefense R&D, examples include policymakers' general concerns about the adequacy of the science, technology, and engineering base (e.g., the creation of the NSF), evolving public health concerns (e.g., Nixon's 'war on cancer'), multinational scientific efforts (e.g., Human Genome Project), certain geopolitical events (e.g., Sputnik and the creation of NASA), or initiatives with mixed diplomatic/scientific objectives (e.g., International Space Station).

Figure 5 shows the changes in nondefense (left) and defense (right) appropriations, expressed in 2012 dollars per capita for ease of comparison across the sample period. The blue

Figure 5: Changes in Nondefense and Defense R&D Appropriations

(a) Nondefense Agencies (b) Defense Agencies




*Notes*: Nondefense agencies include NASA, NIH, NSF, and the nondefense functions of DOE. Defense agencies include DOD and national security functions of DOE. Nominal amounts are converted to real dollars using NIPA price indices for federal nondefense/defense investment in intellectual property. *Source*: Authors' calculations based on the *Budget of the U.S. Government*, Fieldhouse and Mertens (2023).

bars show changes that we classify as exogenous, and the red bars show changes classified as endogenous or those too small to classify (see Appendix C for the figures for each individual agency). A key result from the narrative analysis is that relatively few of the significant changes in R&D appropriations are classified as endogenous to cyclical factors, especially when compared to similar analyses of changes in monetary policy (Romer and Romer 1989), tax policy (Romer and Romer 2010), or credit policies (Fieldhouse et al. 2018). This is perhaps not too surprising in hindsight, as R&D funding makes up a small part of the federal budget, primarily benefits high-skill workers in the short run, and is unlikely to produce significant macroeconomic effects within the electoral cycle. As endogenous policy changes are relatively few and, at times, go in opposite directions in similar macroeconomic circumstances, the classification into endogenous and exogenous policy changes will turn out to be relatively unimportant for the quantitative estimates in this paper. This is, of course, impossible to verify ex-ante without actually conducting the narrative analysis.

## III. The Dynamic Effects of Changes in R&D Appropriations

### A. Empirical Methodology

The first part of our analysis consists of estimating impulse responses of productivity and government R&D capital to unanticipated changes—or 'shocks'—to defense (D) and nondefense (ND) R&D appropriations. To measure these shocks, we first define $z_t^i = \Delta a_t^{exo,i}/K_{t-4}^i$ for $i = D, ND$, where $\Delta a_t^{exo,i}$ are the narrative measures of exogenous real changes in appropriations in quarter $t$ (as shown on a per capita basis in blue in Figure 5) and $K_t^i$ is the stock of R&D capital in category $i$. We scale the changes in R&D appropriations by the capital stock as we are interested in elasticities to government R&D capital. To avoid introducing endogeneity, we scale by the one-year lagged capital stocks, although this matters very little for the results. To estimate impulse responses to unanticipated changes in $z_t^i$, we use Jordà (2005) local projections that include $z_t^i$ as well as a set of predetermined control variables. Given the likely significant delays between an increase in congressional appropriations for R&D, actual outlays for R&D, and any eventual technological improvements as a result of those outlays, we estimate responses at forecast horizons $h = 0, ..., H - 1$ of up to 15 years ($H = 60$ quarters). Unless mentioned otherwise, the estimation sample consists of 74 years of quarterly observations from 1948Q1 through 2021Q4.

In our baseline specification, the impulse response of an outcome variable $y_t$ at horizon $h$ is the OLS estimate of $\gamma_h$ in a direct forecasting regression of $y_{t+h}$ on $z_t^i$:

$$(1) \quad \sum_{j=0}^{3}\left(\frac{1}{4} \times y_{t+h-j}\right) \;=\; c_h + \gamma_h z_t^i + \sum_{j=1}^{p} \beta_h^j \ln(a_{t-j}^i) + \sum_{j=1}^{p} \delta_h^j y_{t-j} + \sum_{j=1}^{p} \zeta_h^{j\prime} x_{t-j} + v_{t,h}$$

where $\ln(a_{t-j}^i)$, $y_{t-j}$, and $x_{t-j}$ for $j = 1, ..., p$ are the predetermined controls and $v_{t,h}$ is a residual at forecast horizon $h$.

Because the changes in R&D appropriations are serially correlated, as seen in Figure 5, we include $p = 4$ quarterly lags of $\ln(a_t^i)$, where $a_t^i$ is the cumulative sum of all past changes in real R&D appropriations in category $i$. Including lags of $\ln(a_t^i)$ rather than $z_t^i$ provides more information about past R&D policies, and adding lags of $z_t^i$ has little effect on the results. We also always include $p = 4$ lags of the outcome variable $y_t$ in all specifications. In addition to the lags of appropriations and the outcome variable, (1) includes lags of a number of additional control variables, $x_t$. As is well known, including lagged predictors can sharpen inference on the impulse response estimates by reducing the variance of the residuals, $v_{t,h}$. Adding a suitable set of lagged controls can also help eliminate past influences on the outcome variable that may be correlated with the regressor of interest and otherwise would lead to endogeneity bias. One of the controls included in $x_t$ is a cyclical indicator to eliminate any remaining cyclical sources of endogeneity in the narrative measures. Specifically, we include the capacity utilization rate from Fernald (2012), which captures variation in both labor effort and the workweek of capital and is strongly correlated with other coincident cyclical indicators.

As discussed earlier, the narrative classification and cyclical controls aim to address the short-run sources of endogeneity that are typically of greatest concern in the identification of fiscal shocks. However, it is not clear that they address potential longer-run sources of policy endogeneity. While relatively few policy changes appear driven by short-term macroeconomic factors, our narrative analysis does not rule out that R&D policies respond systematically to slower-moving productivity, demographic, or other secular trends. A related possibility is that policy responds to the arrival of new ideas and nascent technologies that, even without government involvement, are anticipated to raise productivity growth.

To address concerns about longer-term sources of endogeneity, the baseline specification includes five additional controls to remove predictable variation in TFP and other outcome variables of interest. First, we always include lags of utilization-adjusted TFP (in log levels) in the control set. Next, we include real government and business-sector R&D capital (both in log-levels) in $x_t$. Including R&D capital, rather than just recent R&D expenditures, is preferable because of the potentially long delays between expenditures and actual improvements in productivity. We further include an average of the cumulative real stock market return for the high-tech, manufacturing, and health industries as a forward-looking indicator of innovation and productivity growth.[8] The final element in the baseline control set $x_t$ is the defense spending news variable of Ramey and Zubairy (2018), which we include to remove additional predictable variation in public R&D, see Antolin-Diaz and Surico (2024).

Two other features of our baseline specification in (1) warrant further discussion: First, the dependent variable is a four-quarter backward moving average of the outcome variable, $\sum_{j=0}^{3} \left(\frac{1}{4} \times y_{t+h-j}\right)$, rather than its quarterly observation $y_{t+h}$. The estimates, therefore, measure the response of the four-quarter average of the outcome variable rather than of its quarterly value. The averaging is numerically equivalent to using $y_{t+h}$ as the left-hand side variable and subsequently taking the moving average of the estimated impulse response coefficients, which smooths some of the quarterly noise in the estimates.

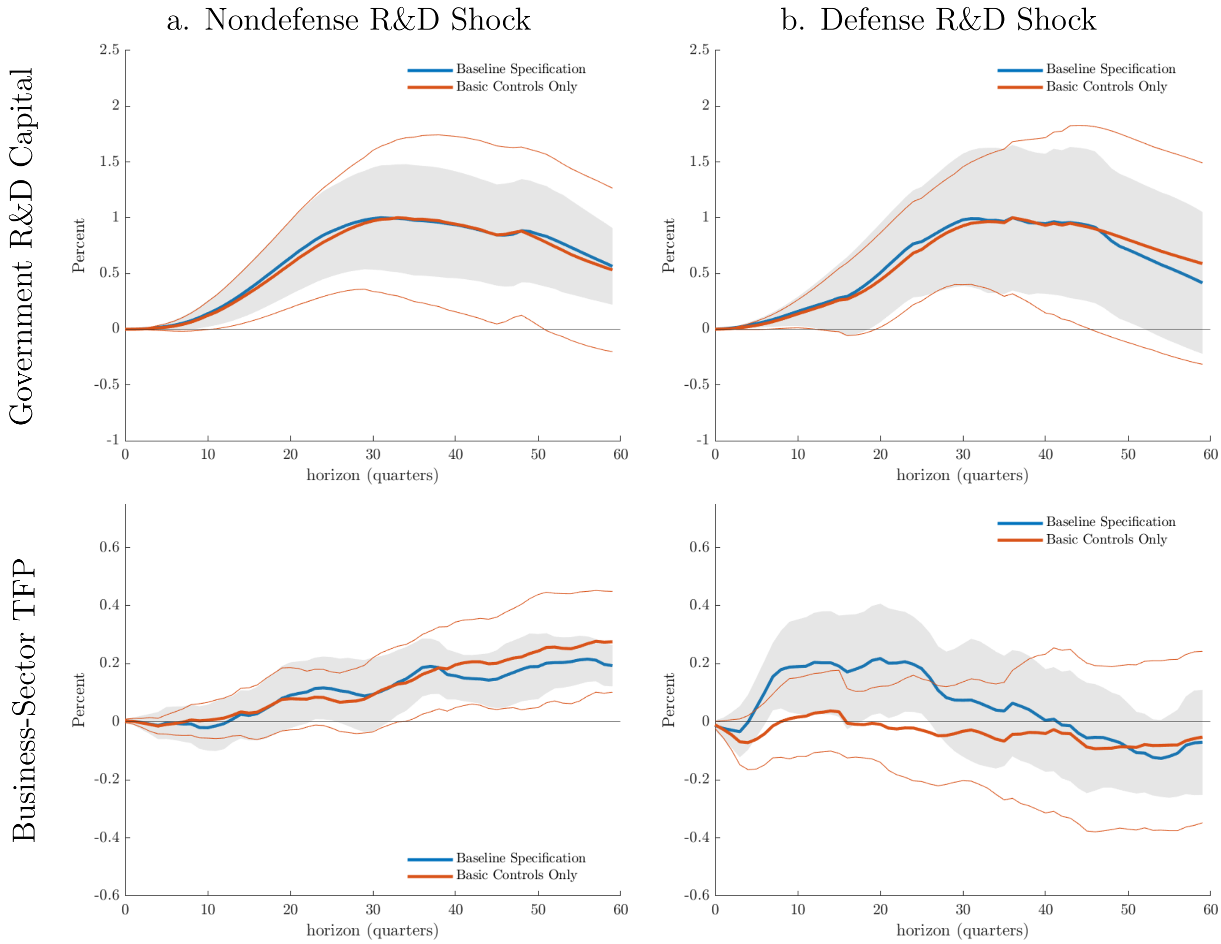
Second, we estimate responses to shocks to nondefense and defense appropriations by simply repeating the baseline specification using $\{z_t^{ND}, ln(a_{t-j}^{ND})\}$ and $\{z_t^{D}, ln(a_{t-j}^{D})\}$, respectively. However, the sample correlation between $z_t^{ND}$ and $z_t^{D}$ is 0.31, which suggests that policy decisions regarding nondefense and defense appropriations are likely not independent. We discuss later how a shock to one category impacts the other, and how additionally controlling for the other component has no major effects on the impulse response estimates.

### B. Government R&D and TFP After Shocks to R&D Appropriations

Figure 6 presents the impulse responses of government R&D capital and TFP to appropriations shocks based on the estimates of $\{\gamma_h\}_{h=0}^{H-1}$ in the local projections in (1). Besides the results from the baseline specification (blue), each panel also shows results for a specification (red) with only a basic set of controls: lags of the outcome variable and cumulated appropriations, therefore excluding the additional controls in $x_t$ described above. The comparison between the baseline and basic-controls-only specifications allows an assessment of the role of including the additional controls. For ease of comparison with the production

[8] Several studies have shown that stock market returns are predictive of output growth and TFP at longer forecast horizons, see e.g., Fama (1990) or Beaudry and Portier (2006). The natural explanation is that new ideas and research opportunities are reflected in stock market valuations relatively quickly and well ahead of the eventual productivity improvements. Indeed, Kogan et al. (2017) document evidence of immediate stock market reactions to patent grants.

Figure 6: Government R&D Capital and TFP Following an Increase in R&D Appropriations




*Notes*: Estimates based on (1) using the narrative measure of changes in federal nondefense (left panel) and defense (right panel) R&D appropriations. 'Basic Controls Only' includes lags of the outcome variable $y_t$ and appropriations $\ln(a_t^i)$ as controls. 'Baseline' additionally includes lags of the controls in $x_t$ described in the main text. Confidence bands are 95 percent HAR based on Lazarus et al. (2018). Impulses scaled to imply a 1 percent peak increase in government R&D capital. Sample: 1948Q1–2021Q4.

function elasticities presented later, the responses are scaled to imply a one percent peak increase in total government R&D capital. The figure reports 95 percent heteroskedasticity and autocorrelation-robust (HAR) confidence bands recommended by Lazarus et al. (2018).

The top panels in Figure 6 show that both the defense and nondefense R&D appropriations shocks lead to highly persistent hump-shaped increases in government R&D capital. The build-up in R&D capital following both types of shocks is gradual, with peak effects that occur 8 to 10 years after the shocks. The substantial delays in the capital responses show that there are, on average, relatively long lags between a positive shock to congressional appropriations for R&D and eventual outlays. As we show below, the modest declines in R&D capital towards the end of the forecast horizon not only reflect depreciation but also eventual reversals in government R&D spending. In the baseline specification, the increases in government R&D capital after both shocks are highly statistically significant for a wide range of horizons, which indicates that each of the narrative measures is a strong predictor of future government R&D spending. It also means that the R&D spending changes are potentially anticipated well in advance, such that basing identification on variation in appropriations rather than spending is preferable to avoid possible bias due to anticipation effects. The point estimates vary little across the baseline and basic-controls-only specifications, and the main impact of the additional controls is to sharpen inference.

The bottom left panel of Figure 6 shows the estimated response of TFP to a nondefense R&D shock. The key finding is that, after a substantial delay, a positive shock to nondefense appropriations leads to a gradual increase in business-sector TFP. Moreover, the TFP increase becomes highly statistically significant at longer horizons. In the baseline specification, there is initially no significant change in TFP for several years, after which TFP slowly increases to a level that is around 0.2 percent higher by the end of the 15-year horizon. In the specification with basic controls only, the TFP response is somewhat larger, up to around 0.27 percent, at the end of the forecast horizon. Including the additional controls again increases the precision of the estimates, but the TFP response is otherwise similar in shape and significant at longer horizons in both specifications.

The bottom right panel of Figure 6 shows that the TFP response to a defense R&D shock differs meaningfully from the response to a nondefense shock. In contrast to the nondefense shock, a positive defense shock leads to a decline in TFP at longer horizons in both specifications. Overall, the estimates of the TFP response to a defense shock are considerably more uncertain. In the specification with only the basic controls, there is no statistically significant impact on TFP at any horizon. The baseline specification shows evidence of a positive near-term TFP response to a defense shock, with point estimates that are significant between two and eight years. The longer-run decrease in TFP is generally not statistically significant in either specification. Unlike for nondefense R&D, there is, therefore, no evidence that defense R&D has positive TFP spillovers in the longer run, at least not within the 15-year time window that we consider.

A plausible explanation for the differing long-run TFP effects of the two shocks is that, while defense technologies can have civilian applications, their dissemination is likely more limited and slower for strategic reasons. Nondefense R&D work, conversely, is not usually classified, and knowledge spillovers are often deliberately promoted to facilitate the commercialization of new technologies (e.g., through NASA's 'Spinoff' Technology Transfer Program or the NIH's Office of Technology Transfer).[9]

**Robustness** The Online Appendix conducts numerous robustness checks that we summarize here. Appendices D.1 and D.2 show that the TFP responses to both shocks remain very similar if we use all appropriations changes instead of just those classified as exoge-

[9]Technology transfers have also been compelled by legislation, e.g., the Stevenson–Wydler Technology Innovation Act of 1980, Federal Technology Transfer Act of 1986, and National Technology Transfer and Advancement Act of 1995.

nous, or if we control for the lagged appropriations and the contemporaneous value of the narrative measure for the other category to isolate more idiosyncratic variation. Appendix D.3 documents that the positive long-run TFP response to a nondefense shock is robust to numerous further additions to the baseline set of controls, including additional cyclical indicators (unemployment rate and output gap), various fiscal policy indicators (public infrastructure capital, debt, taxes, spending, etc.), financial market indicators (interest rates, credit spreads, and broader stock market indices), and alternative potential predictors of future TFP and R&D spending (labor quality, non-R&D business-sector capital, patents, and the relative price of R&D). Together, these results suggest that endogeneity is likely not as serious a concern for government R&D as it is for broader changes in tax or spending policies. Appendices D.4, D.5, and D.6 further show that the positive long-run TFP response to a nondefense R&D shock is robust to various changes in specification (lag length $p$, including lags of $z_t^i$, a balanced sample for all $h$), using alternative scalings of appropriations to construct the narrative measures (by total government R&D capital or potential output), using alternative impulse response estimators (internal instrument VAR, smooth local projections), and different inference procedures (Ecker-Huber-White, wild bootstrap).

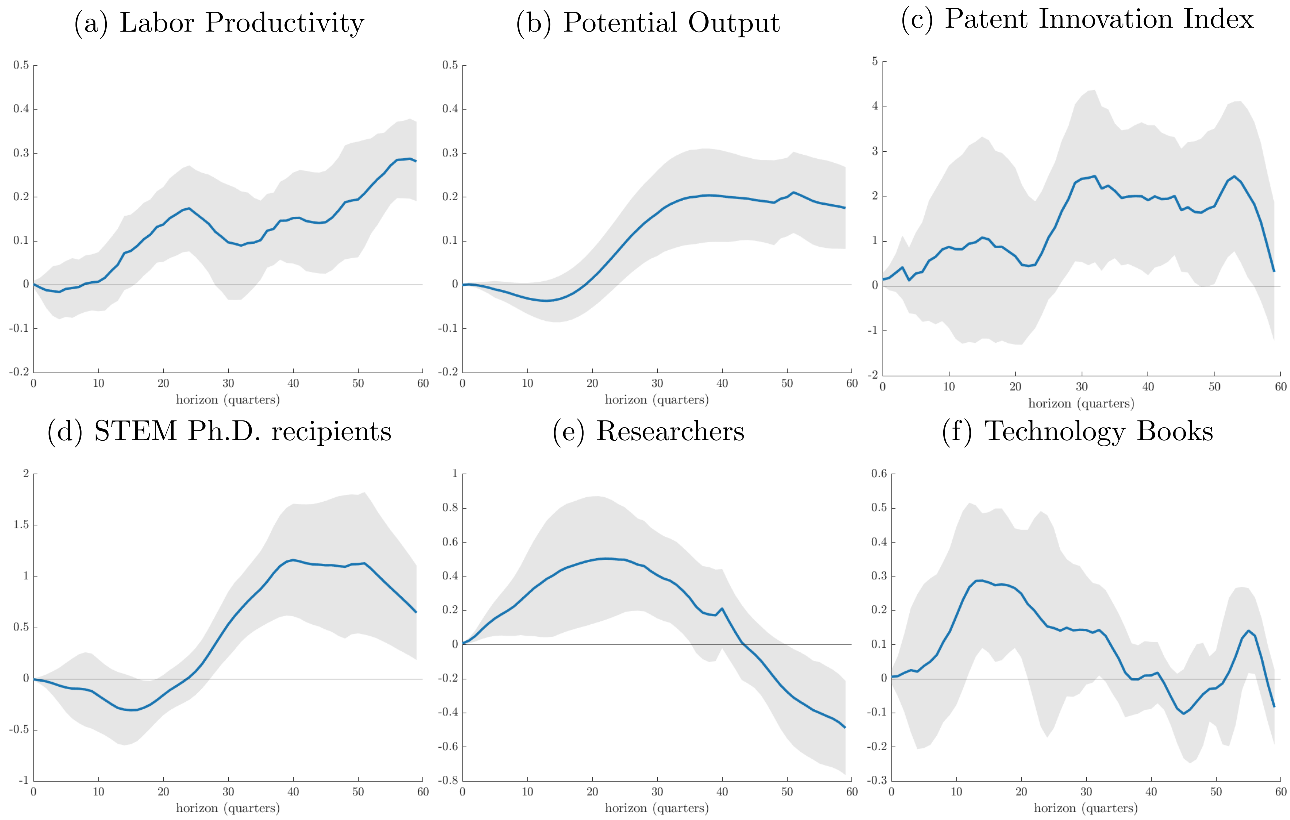
Unless mentioned otherwise, we will maintain the baseline specification—i.e., with lags of the additional controls in $x_t$—in all remaining specifications.

### C. Effects on Other Productivity and Innovation Indicators

Figure 7 reports the responses of several other productivity and innovation indicators to nondefense R&D shocks. For brevity, the corresponding responses to defense R&D shocks are reported in Appendix D.7.

Panel (a) in Figure 7 shows the response of business-sector labor productivity (output per hour). Under certain assumptions, technological change is the only source of long-run variation in labor productivity; see e.g., Galí (1999). The response of labor productivity at longer horizons thus provides an alternative signal of the productivity effects of government R&D. As panel (a) shows, labor productivity initially does not react to a nondefense R&D shock but starts rising after three years and reaches a level that is higher by around 0.25 percent after 15 years. Just as the TFP response to a nondefense shock in Figure 6, the response of labor productivity is highly statistically significant at longer horizons. Appendix D.8 shows that labor input remains essentially flat after a nondefense R&D shock. In contrast, the non-R&D business-sector capital stock rises significantly at longer horizons, with a peak increase of close to 0.2 percent. This pattern of responses is broadly consistent with conventional balanced-growth assumptions in economic models implying that productivity trends have no permanent effect on hours worked per capita. To the extent that the long-run TFP increase is widely anticipated by economic agents, the absence of any short-run response in labor input implies that news about future TFP from changes in R&D appro-

Figure 7: Impact of a Nondefense R&D Shock on Other Productivity/Innovation Indicators




*Notes*: Estimates based on (1) using the narrative measure of changes in nondefense R&D appropriations. Confidence bands are 95 percent HAR based on Lazarus et al. (2018). Impulses scaled to imply a 1 percent peak increase in government R&D capital. Sample: (a),(b),(d): 1948Q1–2021Q4; (c): 1949Q1–2010Q4; (e): 1951Q1–2019Q4; (f): 1956Q1–1997Q4. See Appendix A for variable definitions.

priations is not a source of fluctuations at business cycle frequencies. Indeed, Appendix D.8 documents that real GDP shows no short-run response to a nondefense shock but simply rises gradually along a trajectory that is very similar to that of labor productivity.

The next panel in Figure 7 shows the response of the CBO's measure of potential GDP, which is an estimate of the economy's maximum sustainable output consistent with stable inflation. TFP is a key determinant of the level of potential output, in addition to the levels of labor and capital inputs being utilized at sustainable rates. Similar to the responses of TFP and labor productivity, panel (b) in Figure 7 shows that there is no effect on potential output for the first five or six years after a nondefense shock. In the long run, there is a gradual and significant increase in potential output, which expands by about 0.2 percent after 8 to 15 years. With no response of labor input and non-R&D business-sector capital increasing by around 0.2 percent, the long-run rise in potential output appears primarily driven by the increase in TFP.

Patent data are a widely used alternative to productivity measures for evaluating the pace of technological innovation across time; see, e.g., Kogan et al. (2017), Miranda-Agrippino et al. (2022), and Kelly et al. (2021), among others. Panel (c) in Figure 7 shows the impact of a nondefense R&D shock on the patent-based innovation index of Kogan et al. (2017), using the quarterly version constructed by Cascaldi-Garcia and Vukotić (2022) available through 2010Q4. This index weights new patent grants by stock market reactions to account for their economic value. As seen in panel (c), the patent innovation index rises by up to 2 percent after a positive shock to nondefense R&D appropriations, an increase that is significant at horizons between 7 and 13 years. The rise in patent grants with economic value starts ahead of the increase in TFP seen in Figure 6 and fades near the end of the forecast horizon. The timing of the response appears roughly consistent with increased government funding for nondefense R&D leading to more patents with economic value followed by improvements in business-sector productivity.

The bottom row of Figure 7 shows responses of several other measures of research activity. Because these measures are only available annually, we construct quarterly versions of these annual series by linear interpolation. Panel (d) first depicts the estimated responses of the (log) number of new Ph.D. recipients in STEM fields to a positive nondefense R&D shock. The response shows a statistically significant increase in new STEM Ph.D. recipients at horizons above seven or eight years, a delay that is consistent with the average length of a Ph.D. after allowing for some additional implementation lags. The increase is persistent over longer horizons, with a peak rise in new STEM doctoral degrees of more than one percent after roughly 10 years.

The next panel considers the (log) number of researchers, i.e., the number of full-time equivalent workers engaged in R&D, based on data from the OECD and Bloom et al. (2020). As the panel shows, a nondefense R&D shock leads to a gradual increase in the number of researchers by up to 0.5 percent approximately three- to eight years after the increase in appropriations. Over longer horizons, the number of researchers first returns back to prior levels and then declines at the end of the 15-year horizon. The long-run decline likely reflects the eventual reversal of the increase in government R&D funding.

The last panel in Figure 7 shows the response of an index of new technology book publications, a measure of innovation constructed by Alexopoulos (2011). While the available sample for this series is much shorter (1956 through 1997), there is evidence of a significant increase in new technology books at horizons of three to eight years. As was the case for the innovation index in panel (c) and the number of researchers in panel (e), the effect on the number of technology publications is transitory and occurs ahead of the TFP response.

The evidence in Figure 7 indicates that a nondefense R&D shock leads to increases in both inputs (researchers and STEM scientists) and outputs (patents, technology books) of the knowledge production function. Both in direction and timing, the responses appear consistent with the simplest explanation of the delayed increase in TFP in Figure 6, which is that government funding for research directly leads to innovations that prove valuable in private production.

Appendix D.7 shows that, in contrast to a nondefense shock, a positive shock to defense R&D does not lead to the same consistent increases in the various productivity and innovation indicators, reinforcing our earlier conclusion that a positive shock to defense appropriations does not appear to have the same positive long-run spillovers on business-sector productivity as nondefense R&D, at least not within a similar time frame.

### D. A Closer Look at the Response of R&D Investment

To gain a better understanding of the nature of both types of R&D shocks, we next take a closer look at the responses of R&D investment spending flows using the underlying NCSES survey data. Specifically, we study how the appropriations shocks affect government R&D spending by type, performer, and funding agency using the series in Figures 3 and 4.

To estimate decompositions of the spending changes, we use the following Törnqvist index approximation of the log change in total real R&D investment, $I_t^{tot}$,

$$\Delta \ln I_t^{tot} \approx \sum_j \frac{s_t^j + s_{t-1}^j}{2} \Delta \ln I_t^j \tag{2}$$

where $I^j$ is gross R&D investment in category $j$ in constant dollars and $s^j$ denotes the nominal expenditure share of category $j$ in total R&D investment ($s^j = I^{n,j}/I^{n,tot}$, where $I^{n,j}$ is gross R&D investment in category $j$ in current dollars). To obtain the individual contributions of each category, we estimate the cumulative impulse response of each of the terms of the summation in (2) using the baseline specification in (1). Because the NCSES survey data are only available for fiscal years, we convert the series to calendar years and use linear interpolation to obtain quarterly spending shares. We then apply these shares to the BEA expenditures to construct quarterly series for all the subcategories that are consistent with the NIPA totals. The impulses are scaled such that the peak increase in total spending on nondefense (left panels) or defense (right panels) R&D is one real dollar. The resulting estimates can thus be interpreted as the real dollar changes in spending in category $j$ given a peak increase in nondefense (or defense) spending of one dollar.

The first two rows in Figure 8 show the responses of total government R&D spending, together with the decompositions by type and performing sector. As the figure shows, both shocks lead to a gradual build-up in total R&D spending flows and partial spending reversals in the longer run. R&D spending after a nondefense shock changes little for the first six quarters or so, rising slowly afterward to a peak after about six years and subsequently declining gradually. After about 10 to 12 years, there is a reversal in spending that lasts until the end of the forecast horizon. The response to a defense shock is similar, except that the rise in spending occurs somewhat more quickly.

The decomposition in the first row of Figure 8 shows that both shocks lead to increases in all types of R&D investment during the boom phase. However, the nondefense shock leads to a larger increase in basic and applied research (up to 32 cents and 20 cents, respectively).

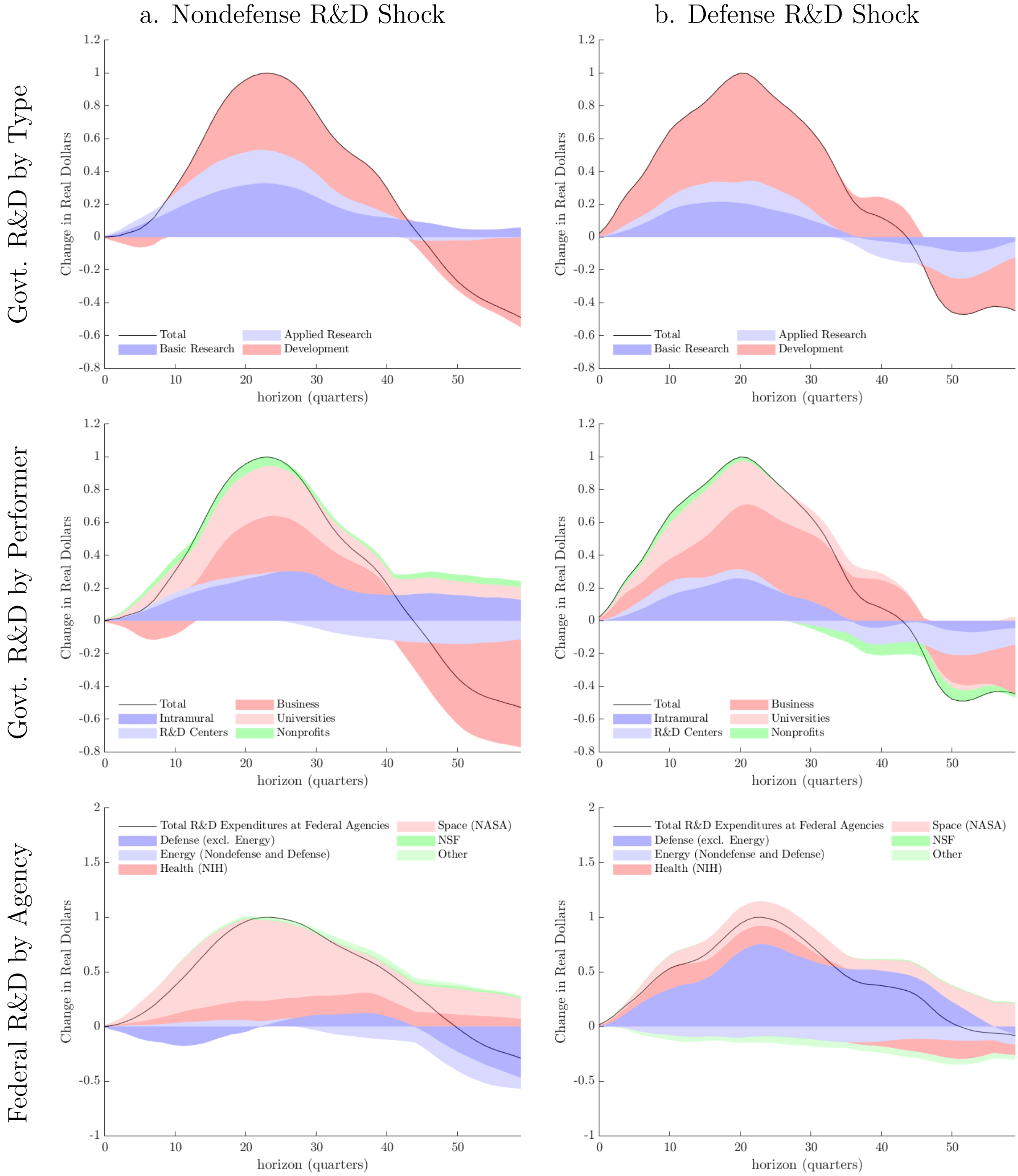
A defense shock instead is tilted more towards spending on development (up to 66 cents). For the nondefense shock, the eventual reversal in spending comes predominantly from development, while funding for basic research remains elevated throughout. For the defense shock, the spending reversal instead occurs in all three types of R&D. As mentioned earlier, Akcigit et al. (2020) argue that basic research generates greater knowledge spillovers than non-basic research. Beyond national security prerogatives limiting knowledge spillovers from defense activities, the larger and more persistent impact of the nondefense shock on basic research may thus contribute to the difference between the long-run TFP responses to defense and nondefense shocks in Figure 6.

As shown earlier in Figure 4, most government R&D spending funds activity that is not performed at federal agencies but at private businesses, public-private R&D centers, or universities. The decomposition in the second row of Figure 8 shows that this is also the case for the spending increases induced by the shocks. A nondefense shock that increases total government R&D investment by up to one dollar raises intramural spending by at most 30 cents, while a defense shock raises intramural spending by at most 26 cents. In both cases, the bulk of the spending increase is funding research conducted by private businesses or universities. For the nondefense shock, the eventual spending reversal is driven almost entirely by decreases in funding for businesses and R&D centers; the increase in funding for research at universities and government agencies is instead more persistent, which likely mirrors the persistent impact of the nondefense shock on funding for basic research. For the defense shock, in contrast, the reversal in spending affects R&D funding for all performers.

The final row in Figure 8 shows a decomposition of the response of federal R&D spending across the main federal funding agencies. As the left panel shows, a nondefense shock leads to persistent increases in funding by NASA, NIH, NSF and other nondefense agencies. Quantitatively, the increase in spending by NASA is by far the largest in size, although NIH funding also sees a meaningful and persistent increase. The increase in nondefense R&D spending overall does not lead to major changes in R&D funding by DOD or DOE, although there is some evidence of crowding out of DOD outlays towards the end of the forecast horizon. Unsurprisingly, the bottom right panel of Figure 8 shows that a positive defense R&D shock mainly leads to DOD spending increases, with smaller increases in funding for health research and NASA. Overall, there is little evidence that a defense shock has large crowding-out effects on R&D funding by nondefense agencies.

**The Role of NASA** The decomposition by federal agency shows that, in dollar terms, the nondefense shock induces large changes in R&D funding for NASA. This finding raises the concern that changes in appropriations for NASA, especially during the agency's rapid growth during the space race, play an outsized role in driving our results. Appendix D.9 considers responses to a nondefense shock based on a narrative measure that excludes all appropriations for NASA, and a measure that only includes NASA appropriations. In both

Figure 8: Response of R&D Spending by Type, Performer and Agency




*Notes*: Estimates based on (1) using the narrative measure of changes in federal nondefense (left panel) and defense (right panel) R&D appropriations. Impulses scaled to imply a unit peak increase in government R&D expenditures (row 1 and 2) or federal R&D expenditures (row 3). See notes in Figures 3 and 4 for data sources. Quarterly values are obtained by interpolation of annual data. Real variables based on the NIPA deflator for government intellectual property (R&D and software). Sample: 1954Q1–2021Q1.

cases, TFP rises by a similar amount as in the baseline, peaking around 0.20, as in the bottom left panel of Figure 6. The TFP increases are also statistically significant at longer horizons, both after excluding NASA appropriations or using only NASA appropriations.

That said, omitting NASA appropriations increases the uncertainty around the estimates considerably, as they contain some of the largest changes within the narrative nondefense measure. Interestingly, the effects of a non-NASA appropriations shock occur somewhat more quickly, whereas the effects of a NASA shock are more delayed relative to the baseline.

Given the importance of NASA appropriations as a source of identifying variation, when estimating production function elasticities and rates of return in Sections 4 and 5 below, we will consistently verify the sensitivity of the results using alternative specifications that omit all NASA appropriations from the narrative measure.

### E. Indirect Channels for Long-Run TFP Spillovers

The evidence presented so far is consistent with a significant direct effect of nondefense government R&D on the level of innovative activity with spillovers on business-sector productivity. However, the long-run TFP responses in Figure 6 are potentially also shaped by additional indirect effects. For example, the appropriations shocks may affect other long-run determinants of productivity growth, such as R&D funding by the private sector or resources allocated to public infrastructure, which could have independent spillovers on business-sector productivity. We next explore the importance of these indirect channels.

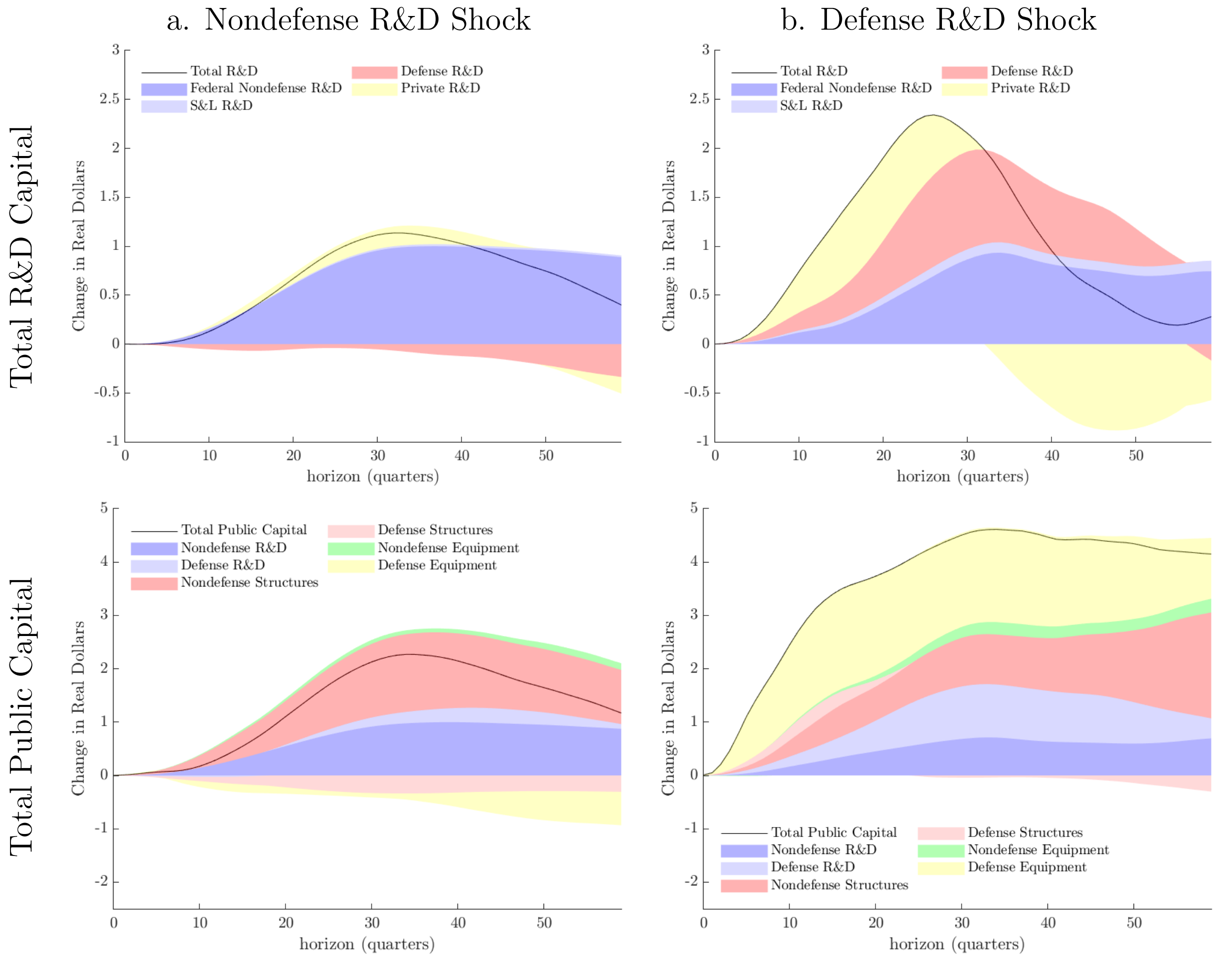
We first investigate how changes in appropriations affect total R&D capital in the economy (private and public). The top row in Figure 9 presents a decomposition of the impulse response of total R&D capital into the individual contributions of each funding sector. These contributions are estimated as in the previous section, using the following approximation of the log change in total R&D capital, $K_t^{tot}$,

$$\Delta \ln K_t^{tot} \approx \sum_j \frac{s_t^j + s_{t-1}^j}{2} \Delta \ln K_t^j \tag{3}$$

where $K^j$ is R&D capital of category $j$ in constant dollars and $s^j$ denotes the nominal share of capital of category $j$ in total R&D capital ($s^j = K^{n,j}/K^{n,tot}$, where $K^{n,j}$ is capital in category $j$ in current dollars). The four main funders of total R&D are (i) federal defense agencies, (ii) federal nondefense agencies, (iii) state and local (S&L) governments, and (iv) the private sector. The contributions of each funding category are the cumulative impulse response of the individual terms in (3) estimated with the baseline specification in (1). The impulses are scaled such that the peak increase in federal nondefense R&D capital (left panel) or defense capital (right panel) is one real dollar. The resulting estimates can be interpreted as the real dollar change in capital in category $j$ given a peak increase in nondefense (or defense) R&D capital of one dollar.

The top left panel in Figure 9 shows that the nondefense shock primarily affects federal nondefense R&D capital, with some mild crowding out of defense R&D capital by up to 34 cents after 15 years. A positive defense shock increases defense R&D capital but also

Figure 9: Total R&D and Public Capital Following an Increase in R&D Appropriations




*Notes*: Estimates based on (1) using the narrative measure of changes in federal nondefense (left panel) and defense (right panel) R&D appropriations. R&D capital includes software. Impulses scaled to imply a unit peak increase in federal nondefense (left) or defense (right) R&D capital. See Appendix A for variable definitions. Sample: 1948Q2–2021Q4.

induces a meaningful increase in nondefense R&D capital (up to 93 cents), as seen in the top right panel. Consistent with the framework in Akcigit et al. (2020) and the evidence in De Lipsis et al. (2022), privately-funded R&D capital increases in response to both shocks. The increases in private R&D capital following a nondefense shock, however, are relatively small, peaking at 20 cents per federally funded dollar. For defense R&D, the peak increase in private R&D capital is larger at 85 cents after around 5 years, but at longer horizons the positive defense shock leads to a decline in private R&D capital of up to 88 cents.[10]

To investigate how R&D appropriations shocks affect other components of the public capital stock, the bottom panels of Figure 9 depict analogous decompositions of the response of total public capital by type. The responses, in this case, are scaled to induce a one-dollar peak increase in government R&D capital in the nondefense (left) or defense (right) category.

[10]Because of data limitations, the possible spillovers on R&D spending in the rest of the world are beyond the scope of this paper. See Moretti et al. (2021) for evidence of crowding-in of defense R&D spending across OECD countries.

The bottom left panel in Figure 9 shows that, in the decomposition of total public capital, there is little evidence of crowding-out of defense R&D capital, even in the longer run. However, the nondefense shock does lead to a broader reallocation from defense to nondefense public capital, with defense structures (e.g., military bases and facilities) and equipment (e.g., ships and aircraft) both declining. In addition, there is a relatively large increase in nondefense structures (e.g., schools and universities, roads, power, water, and sewer systems). For a peak increase in nondefense R&D capital of one dollar, the stock of nondefense structures rises by up to 1.46 dollars after about 8 years. While this increase is much smaller than the average ratio of nondefense structures to R&D capital, it is too large to be explained exclusively by the reclassification of R&D plant expenditures as 'structures' in the national accounts, as these average only 5 percent of non-plant R&D expenditures. Appendix D.10 presents further decompositions showing that about 80 percent of the increase in nondefense structures originates with state and local governments, which finance most nondefense public infrastructure. The growth in state and local structures is broad-based, with the largest increases in education structures (schools and universities), followed by highways and streets, and power, water, and sewer systems. Given the shared funding arrangements for interstate highways, one possibility is that nondefense R&D appropriations are positively correlated with federal transfers for interstate highway spending. However, federal appropriations bills for R&D generally do not provide significant funding for public infrastructure investment via transfers to state and local governments. In Appendix D.10, we further show that federal transfers, if anything, decline in response to a positive nondefense R&D shock. The rise in state and local investment is instead financed initially by debt and later on by increases in tax revenues relative to other expenditures. A more plausible reason why federal R&D appropriations increase investment in education structures is that federal research grants to universities include overhead costs, which can exceed 60% of the total grant and are used to fund facilities, equipment, and administration. Apart from this channel, in our view, the most natural explanation is simply that the rise in economic activity (and the associated increase in business-sector fixed capital documented in Appendix D.8) creates additional needs for public infrastructure investment.

The bottom right panel in Figure 9 shows that a defense R&D shock is also associated with increases in other types of public capital. In general, defense shocks cause smaller changes in the nondefense capital categories, although nondefense structures do increase meaningfully toward the end of the forecast horizon. There is, however, a large and immediate increase in defense equipment (up to 1.88 dollars) and also a smaller increase in defense structures (up to 29 cents). For defense functions, it is easier to point to direct linkages between appropriations for R&D and other military investments. For example, the BEA treats the 'operational systems development' component of DOD's Research, Development, Test, and Evaluation budget accounts as gross investment in equipment, not R&D. More importantly, the annual DOD appropriations bills that fund defense R&D also fund procurement (i.e., equipment), and funding for military hardware development typically leads to subsequent procurement of newly operational weapons systems. In addition, the same geopolitical events that motivate significant increases in defense R&D are likely to also motivate other military investments that may not be fully predicted by the Ramey and Zubairy (2018) military news variable in the set of controls.

The main conclusions regarding the indirect channels for long-run TFP spillovers are the following: First, the impact of a shock to nondefense R&D on defense R&D is relatively small, whereas a defense R&D shock is followed by a sizable increase in nondefense R&D. Second, positive shocks to R&D appropriations lead to higher private spending on R&D, such that private and government-funded R&D capital appear to be complements rather than substitutes. That said, the increases in privately funded R&D capital after a nondefense shock are relatively small, and they are larger but transitory after a defense shock, before declining at longer horizons. We also reiterate that the bulk of the increase in public R&D funding is for research activity performed extramurally; see the second row in Figure 8. Despite the relatively small increase in private R&D capital after a nondefense shock, much of the additional innovative activity following an increase in appropriations occurs outside of the government sector. Finally, the responses of public infrastructure capital are sizable. There is little evidence in the literature that defense equipment and structures have any effect on private sector productivity, and the convention is to assume no effect, e.g., CBO (2016). However, there is widespread evidence for productivity spillovers of nondefense public infrastructure. Since the response of nondefense public infrastructure is potentially important in determining the long-run TFP responses, we will account for those indirect spillover effects in the analysis below.

## IV. The Production Function Elasticity of Government R&D

Figure 6 shows that a nondefense appropriations shock raising government R&D capital by one percent increases TFP by around 0.2 percent in the longer run. The results in the previous section suggest that indirect effects may contribute meaningfully to the TFP response, in particular through the impact on nondefense public infrastructure. To isolate the direct spillover effects on business-sector productivity, in this section, we structurally estimate the aggregate production function elasticity of government R&D capital.

### A. Empirical Methodology and Identification Assumptions

The starting point is the following aggregate production function for quarterly aggregate output growth in the business sector,

$$\Delta f_t = \alpha_t' \Delta m_t + \eta_t \Delta q_t + \phi_t \Delta k_t + \Delta \nu_t \tag{4}$$

where $f_t$ is the log of real business-sector output, the vector $m_t$ collects all business-sector capital (including R&D) and labor inputs in logs, $q_t$ is the log of the public infrastructure capital stock, $k_t$ is the log of government R&D capital, and $\nu_t$ is technological progress after accounting for growth in both types of public capital, $q_t$ and $k_t$. The parameters in $\alpha_t$ are the production function elasticities of private inputs, $\eta_t$ is the elasticity of public infrastructure, and $\phi_t$ is the elasticity of government R&D capital. Following convention in the literature, we assume that (non-R&D) defense capital does not generate any TFP spillovers.

Defining $\Delta tfp_t = \Delta f_t - \alpha_t' \Delta m_t + \epsilon_t$, and assuming constant elasticities with respect to public capital, (4) can be rewritten as

$$\Delta tfp_t \quad = \quad \eta \Delta q_t + \phi \Delta k_t + \Delta w_t \quad , \quad \Delta w_t = \Delta \nu_t + \epsilon_t \quad , \quad E[\Delta w_t] = 0 \tag{5}$$

where $\Delta tfp_t$ is the utilization-adjusted measure of business-sector TFP growth constructed by Fernald (2012) and $\epsilon_t$ is measurement error. The unobserved residual term $\Delta w_t$ consists of the productivity growth term $\Delta \nu_t$, as well as any discrepancy $\epsilon_t$ between measured TFP and actual productivity growth. Apart from measurement errors in $\Delta f_t$ and $\Delta m_t$, the discrepancy between measured and actual productivity growth could be due to the mismeasurement of the elasticities in $\alpha_t$. As explained in Fernald (2012), the identification of $\alpha_t$—which in practice is based on factor cost shares—relies on theoretical assumptions that may not hold in reality, for instance, the absence of markups. As a result, $\epsilon_t$ cannot generally be treated as classical measurement error, as it potentially also contains the influence of all determinants of private factor inputs, including shocks to government R&D. The other endogeneity concern is that movements in the residual productivity term $\Delta \nu_t$ are correlated with government investment.

Our strategy to address endogeneity relies on two steps. First, we treat $\eta$ as a known constant and estimate $\phi$ across a range of values of $\eta$ consistent with the empirical literature. A recent survey by Ramey (2021) establishes a plausible range of 0.065 to 0.12 for $\eta$. We use these endpoints to estimate a corresponding range for $\phi$, and we also consider the intermediate value of $\eta = 0.08$, which is the value that the CBO currently uses to quantify the impact of public infrastructure (CBO 2021). Treating $\eta$ as known, we define $\Delta \widetilde{tfp}_t \equiv \Delta tfp_t - \eta \Delta q_t$, i.e., the growth in measured TFP adjusted for the productivity effects of public infrastructure capital. Substituting into (5), this definition leads to the structural estimation equation,

$$\Delta \widetilde{tfp}_t \quad = \quad \phi \Delta k_t + \Delta w_t \tag{6}$$

where in general $E[\Delta k_t \Delta w_t] \neq 0$ such that endogeneity remains a concern.

The second step in our identification strategy is to estimate $\phi$ in (6) using the SP-IV estimator of Lewis and Mertens (2023). The SP-IV estimator is a GMM estimator with an intuitive closed-form solution as the OLS estimate in a regression of estimated impulse responses to shocks that are uncorrelated with the structural error, see also Appendix E.1. In our application, we use the responses to R&D appropriations shocks discussed earlier.[11] Note that, in contrast to the local projections used for estimating the impulse responses, the functional form in (6) makes very specific assumptions about the lags between R&D spending and the productivity effects. Appendix E.1 shows that these assumptions, in fact, align very well with the impulse response estimates of $\widetilde{tfp}_t$ and $k_t$, which follow very similar hump-shapes that are approximately proportional.

To understand the identifying moments in the GMM problem that generates the SP-IV estimator, let $\Omega_{t-1} \equiv \{\ln a^i_{t-j}, y_{t-j}, x_{t-j}\}_{j=1}^p$ define the full set of lagged controls included in the local projections in (1). Letting $z_t$ denote the $N_z \times 1$ vector containing the $N_z$ narrative measures used for identification, the $HN_z$ moment conditions that identify $\phi$ are

$$(7) \qquad E\big[w_t^\perp(h) z_t^\perp\big] = 0 \; ; \; h = 0, \dots, H-1 \quad , \; w_t^\perp(h) \equiv \widetilde{tfp}_t^\perp(h) - \phi k_t^\perp(h)$$

where $z_t^\perp$ is the one-step ahead forecast error from the linear projection of $z_t$ on $\Omega_{t-1}$ and $\widetilde{tfp}_t^\perp(h)$ and $k_t^\perp(h)$ are the $h+1$-step ahead forecast errors from linear projection of $\widetilde{tfp}_{t+h}$ and $k_{t+h}$ on $\Omega_{t-1}$. Intuitively, the identifying conditions in (7) exploit the fact that, if the structural relationship in (6) holds in the raw data, it also holds across the $h+1$-step ahead forecast errors after projection on $\Omega_{t-1}$ for any forecast horizon $h$. The key exogeneity assumption in (7) is that, after projection on $\Omega_{t-1}$, the period $t$ innovations in the narrative measures, $z_t^\perp$, are uncorrelated with the ex-post deviations $w_t^\perp(h)$ from the structural relationship across the period $t$ forecast errors at all horizons $h = 0, ..., H-1$.

The conditional forecast errors $w_t^\perp(h)$ arise either because of accumulated technological progress $\Delta\nu$ between period $t$ and $t+h$ that is not predicted by the projection on $\Omega_{t-1}$, or because of accumulated measurement error $\epsilon$ in measured TFP between period $t$ and $t+h$ that is unpredicted by projection on $\Omega_{t-1}$. The first part of the exogeneity requirement is a zero correlation between $z_t^\perp$ and all sources of unpredicted productivity growth between $t$ and $t+H-1$ that are not driven by the accumulation of government R&D capital. Changes in appropriations in quarter $t$ are plausibly uncorrelated with *future* realizations of unanticipated technology shocks in quarters $t+h > t$. The narrative classification step is intended to preclude any *contemporaneous* nonzero correlation between R&D appropriations and technology shocks at $h = 0$. In addition, the typical recognition and legislative lags in fiscal policy arguably make any systematic policy reaction to technology shocks within the same quarter unlikely. Finally, we assume that conditioning on the variables in $\Omega_{t-1}$ suffices to remove any joint influences of *past* shocks (realized prior to quarter $t$) on $z_t$ and future productivity growth.

[11] One minor difference is that the impulses underlying the SP-IV estimator are estimated in balanced samples rather than iteratively, as is required for the inference formulas developed in Lewis and Mertens (2023). Appendix D.4 shows that the impulse response estimates are very similar in the balanced sample to those shown in Figure 6.

The second part of the exogeneity requirement is that $z_t^\perp$ is uncorrelated with any unpredicted accumulated measurement error in TFP across the forecast horizon. If the measurement error in TFP is strictly exogenous, the identifying conditions in (7) remain perfectly valid. If the error is the result of mismeasurement of the elasticities of private factor inputs, then $w_t^\perp(h)$ is generally a function of any shock that causes changes in the factor inputs for which the elasticities are mismeasured. In that case, we appeal to the same arguments as above to motivate the assumption that $z_t^\perp$ is not correlated with other shocks: non-causal correlations with future non-technology shocks are implausible, the narrative classification and policy lags eliminate any contemporaneous correlations with non-technology shocks, and the control set $\Omega_{t-1}$ removes any confounding influences of past non-technology shocks.

The same arguments do not apply, however, to the R&D appropriations shocks themselves. If appropriations shocks cause meaningful changes in private factor inputs, and these changes are not properly accounted for in the measurement of TFP, then (7) would not necessarily hold, and the SP-IV estimate of $\phi$ would be potentially biased. Fortunately, the estimated impulse responses of business-sector labor and non-R&D capital inputs to R&D appropriations shocks, reported in Appendix D.8, imply that any errors in the production function elasticities for these factor inputs would have to be very large to introduce a quantitatively significant source of bias.[12] Mismeasurement could be a more serious concern for private R&D capital because of knowledge spillovers, which are not necessarily well captured by the cost share of private R&D. As shown earlier, both R&D shocks lead to increases in private R&D capital. If the methodology in Fernald (2012) underestimates the elasticity of private R&D capital, the estimates of $\phi$ are likely to be biased upward. However, as we discuss in the robustness section below, assuming larger values of the elasticity of private R&D capital has little effect on the estimates of $\phi$, which suggests that any bias is likely very small. Global spillovers through changes in R&D spending abroad are another potential source of bias, but their importance or direction is not immediately obvious.

The estimation equation in (6) does not distinguish between defense and nondefense government R&D capital, whereas the TFP responses in Figure 6 indicate that the spillovers on business-sector productivity are potentially quite different. Moreover, while we saw earlier that a nondefense shock induces little change in defense R&D capital, a defense shock is followed by a sizable increase in nondefense government R&D capital. We, therefore, also consider specifications that include both types of government R&D capital simultaneously, allowing for different elasticities of defense and nondefense government R&D capital. Using the approximation $\Delta k_t \approx s_{ND,t}\Delta k_t^{ND} + (1 - s_{ND,t})\Delta k_t^D$, where $s_{ND,t}$ is the nominal nondefense share of total government R&D capital averaged over $t$ and $t-1$, the estimation

[12]For example, following a nondefense shock that increases government R&D capital by one percent, there is a gradual and statistically significant increase in business-sector non-R&D capital of up to 0.2 percent, see Appendix D.8. Assuming a measured elasticity of non-R&D capital of 0.33, a one-basis-point effect on measured TFP would require a 15 percent error in the capital elasticity ($0.2 \times 0.33 \times 0.15 \approx 0.01$).

equation is adjusted as follows:

$$(8) \qquad \Delta\widetilde{tfp}_t = \phi_{ND}\left(s_{ND,t}\Delta k_t^{ND}\right) + \phi_D(1-s_{ND,t})\Delta k_t^D + \Delta w_t \quad , \quad E[\Delta w_t] = 0$$

This specification assumes production function elasticities to $\Delta k_t^{ND}$ and $\Delta k_t^D$ that scale with $s_{ND,t}$ and $1-s_{ND,t}$, such that $\phi_{ND}$ ($\phi_D$) measures the percent change in TFP for a one percent increase in total government R&D capital that is driven exclusively by an increase in nondefense (defense) R&D capital. The scaling has the advantage that the magnitudes of $\phi_{ND}$ and $\phi_D$ can be compared to the estimates of $\phi$ in the simpler specification in (6). For the purpose of calibrating an aggregate production function with constant elasticities on $\Delta k_t^{ND}$ and $\Delta k_t^D$, the estimates of $\phi_{ND}$ and $\phi_D$ can be multiplied by 0.5, which is approximately the average of $s_{ND,t}$ across the sample. An alternative approach, discussed further below, treats the elasticities to $\Delta k_t^{ND}$ and $\Delta k_t^D$ as constants in the estimation.

When $\phi_{ND} \neq \phi_D$, the estimates of $\phi$ in the simpler specification in (6) are not necessarily consistent for either $\phi_{ND}$ or $\phi_D$. In that case, the response to a nondefense shock only identifies $\phi = \phi_{ND}$ in two situations: either a nondefense shock does not lead to any changes in defense R&D capital, or there are no productivity effects of defense R&D, $\phi_D = 0$. Similarly, the response to a defense shock only identifies $\phi = \phi_D$ if there is no impact on nondefense R&D capital, or else if $\phi_{ND} = 0$. As discussed earlier, the impulse responses to a nondefense shock do not show much impact on defense R&D capital, such that we expect both specifications to provide similar estimates of $\phi^{ND}$. The same is not true for the defense shock, which was followed by sizable increases in nondefense R&D capital. We, therefore, expect estimates of $\phi$ identified with a defense shock to be potentially quite different from those of $\phi^D$.

As is well known, IV estimation can be unreliable when identification is too weak. Applying the diagnostic test of Lewis and Mertens (2023) reveals that weak instruments are a concern in a few of the specifications that we consider below. For this reason, we use the weak-instrument-robust GMM inference methods of Kleibergen (2005), which remain valid regardless of the strength of identification. Other problems can arise when the number of identifying moments is too large (Newey and Windmeijer 2009). Given the high persistence in the impulse response estimates for $\widetilde{tfp}$ and $k$, there is limited additional identifying information in immediately adjacent quarterly horizons. To mitigate potential many-instrument problems, we therefore do not use all horizons for identification, but only those at one-year intervals, at $h = 3, 7, 11, ..., 59$.[13]

[13]Identification is therefore based on 15 moments (rather than 60) for specifications identified with a single impulse response, and 30 moments (rather than 120) for those identified with two. While the application is different, simulation results in Lewis and Mertens (2023) for the estimation of the hybrid New Keynesian Phillips curve indicate that Kleibergen (2005) inference for SP-IV displays only small size distortions in samples of 250 quarters and 20 identifying horizons.

TABLE 1: ESTIMATES OF PRODUCTION FUNCTION ELASTICITIES OF GOVERNMENT R&D CAPITAL

| | Public R&D Measure | Instruments | Intermediate $\eta = 0.08$ $\hat{\phi}/\hat{\phi}_{ND}$ | Intermediate $\eta = 0.08$ $\hat{\phi}/\hat{\phi}_{D}$ | Low $\eta = 0.065$ $\hat{\phi}/\hat{\phi}_{ND}$ | High $\eta = 0.12$ $\hat{\phi}/\hat{\phi}_{ND}$ |
|---|---|---|---|---|---|---|
| [1] | Total | Exo ND | 0.11*** (0.09,0.15) | | 0.11*** (0.09,0.15) | 0.10*** (0.08,0.13) |
| [2] | Total | Exo ND, No NASA | 0.11*** (0.08,0.20) | | 0.12*** (0.08,0.21) | 0.10*** (0.07,0.19) |
| [3] | Total | All ND | 0.10*** (0.09,0.14) | | 0.11*** (0.09,0.15) | 0.09*** (0.07,0.13) |
| [4] | Total | Exo D | | −0.13 (−1.20,0.04) | | |
| [5] | Total | All D | | −0.11 (−1.11,0.05) | | |
| [6] | ND/D | Exo ND | 0.10*** (0.06,0.19) | −0.01 (−0.22,0.39) | 0.11*** (0.06,0.20) | 0.09*** (0.05,0.18) |
| [7] | ND/D | Exo ND/D | 0.10*** (0.04,0.19) | −0.07 (−0.27,0.40) | 0.10*** (0.04,0.19) | 0.09*** (0.03,0.18) |
| [8] | ND/D | Exo ND, No NASA | 0.11 (−2.00†,0.58) | 0.20 (−2.00†,0.69) | 0.11 (−2.00†,0.60) | 0.10 (−2.00†,0.54) |
| [9] | ND/D | All ND | 0.10*** (0.06,0.18) | −0.03 (−0.23,0.35) | 0.10*** (0.06,0.18) | 0.09*** (0.05,0.17) |

*Notes:* Rows [1]-[5]: SP-IV estimates of $\phi$ in (6); rows [6]-[9]: SP-IV estimates of $\phi_{ND}$ and $\phi_D$ in (8). 95 percent weak-instrument-robust intervals based on the KLM statistic of Kleibergen (2005) in parentheses. Test inversion is on a grid with endpoints −2 and 2, † denotes intervals constrained at these endpoints. Subvector inference in rows [6]-[9] uses the projection method. $*$, $**$ and $***$ denote statistical significance at 10, 5 and 1 percent levels, respectively. 'Exo ND/D': exogenous changes in nondefense/defense R&D appropriations. 'All ND/D': all changes in nondefense/defense R&D appropriations. 'No NASA': excluding all NASA appropriations. Sample: 1948Q1–2021Q4.

## B. Estimation Results

Table 1 reports estimates of $\phi$, $\phi_{ND}$, and $\phi_D$ for various specifications, together with 95 percent weak-instrument-robust confidence intervals. The first five rows show estimates of $\phi$ in (6), including only total government R&D capital, whereas the remaining rows show estimates for $\phi_{ND}$ and $\phi_D$ in (8), with nondefense and defense R&D capital stocks included separately. The first two columns report results for TFP adjusted for public infrastructure, $\widetilde{tfp}_t$, using the benchmark value of $\eta = 0.08$. The last two columns show the elasticity estimates based on variation in nondefense R&D capital using the lower and higher values of $\eta = 0.065$ and 0.12, respectively. For brevity, the elasticity estimates based on variation in defense R&D capital for the alternative values of $\eta$ are omitted.

The first row in Table 1 shows estimates based on the impulse responses identified with the narrative measure for nondefense appropriations, $z_t^{ND}$. For $\eta = 0.08$, the point estimate of $\phi$ in (6) based on the impulse responses to a nondefense shock is 0.11. This estimate is highly statistically significant and fairly precisely estimated, with a 95 percent robust confidence interval ranging from 0.09 to 0.15. Assuming a larger elasticity of public infrastructure means that a greater portion of the TFP increase after a nondefense R&D shock in Figure 6 is attributed to the increase in public infrastructure shown in Figure 9, see also Appendix E.1. Consequently, the increase in TFP after adjusting for public infrastructure is smaller when $\eta$ is larger. As expected, the estimates of $\phi$ are decreasing in the assumed value of $\eta$, but in practice, they remain very similar in size across Ramey's (2021) plausible range of values for $\eta \in [0.065, 0.12]$.[14]

Rows [2] and [3] in Table 1 show results based on impulse responses identified with different measures of nondefense R&D appropriations. Row [2] shows the estimates when the narrative nondefense measure omits all changes in appropriations for NASA. While the underlying impulse response estimates are considerably less precise, the elasticity estimates remain highly significant and very close in size to row [1]. Without NASA appropriations as identifying variation, however, the robust confidence intervals become notably wider, and in particular, substantially larger values of $\phi$ cannot be ruled out. Row [3] shows estimates based on responses to all changes in nondefense R&D appropriations, regardless of their narrative classification. The estimates are again very similar to those in row [1], implying that the narrative classification matters little for the identification of $\phi$.

The next two rows in Table 1 report estimates of $\phi$ identified with defense shocks instead of nondefense shocks. Row [4] reports $\hat{\phi} = -0.13$ based on the narrative measure of exogenous changes in defense R&D appropriations $z_t^D$, and row [5] shows that $\hat{\phi} = -0.11$ when using all changes in defense R&D appropriations regardless of their narrative classification. Different from the nondefense shocks, both estimates are negative. The confidence bands are very wide, and none of the estimates are significant at conventional levels. Moreover, pre-testing results based on Lewis and Mertens (2023) fail to reject weak identification, such that the defense-based point estimates likely suffer from substantial small sample bias.

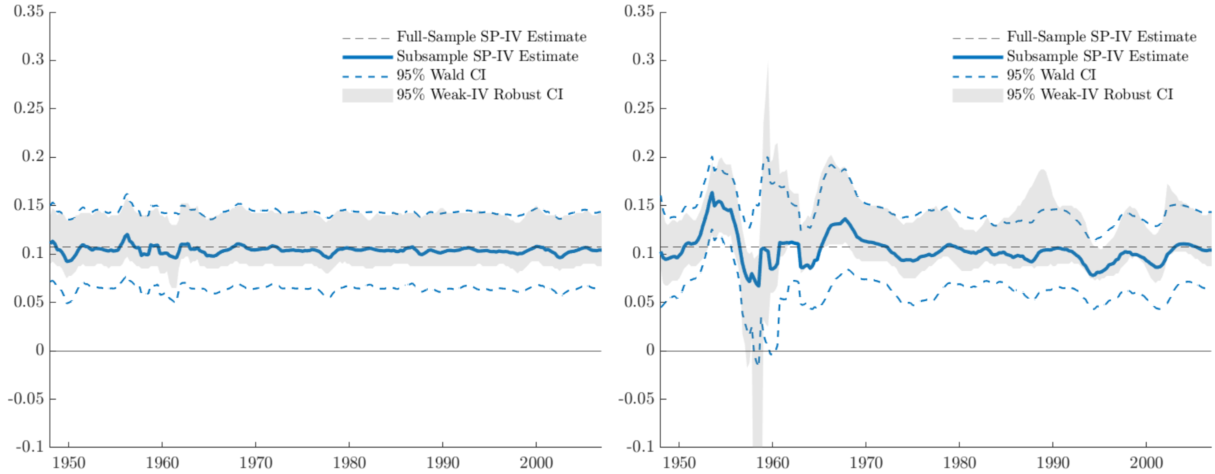
The remaining rows in Table 1 report estimates of $\phi_{ND}$ and $\phi_D$ from the specification in (8) that includes both types of government R&D capital simultaneously, with subvector inference based on the projection method. In row [6], $\phi_{ND}$ and $\phi_D$ are identified jointly using the same nondefense measure as in row [1]. The resulting point estimates of $\phi_{ND}$ are 0.10 for the intermediate value of $\eta$, 0.11 for the low value of $\eta$, and 0.09 for the high value of $\eta$, all of which are statistically significant and close to the corresponding estimates in row [1]. In contrast, the point estimate of $\phi_D$ is small, -0.01, and statistically insignificant.

Rows [7]-[9] in Table 1 provide additional estimates of $\phi_{ND}$ and $\phi_D$ identified with different impulse responses. In row [7], identification is based on both narrative measures, $z_t^{ND}$ and $z_t^D$, simultaneously. Note that, in this case, the non-zero sample correlation between $z_t^{ND}$ and $z_t^D$ is irrelevant for the interpretation of the results. Rows [8] and [9] are instead based on the same narrative measures as in rows [2] and [3], i.e., excluding all NASA appropriations and using all changes in nondefense R&D appropriations, respectively. The estimates of $\phi_D$ across rows [7]-[9] range from -0.07 to 0.20, but are statistically insignificant

[14]The point estimate is $\hat{\phi} = 0.15$ when assuming $\eta = 0$, and $\hat{\phi} = 0.03$ when $\eta = 0.39$. The latter value is that estimated by Aschauer (1989) and is the highest estimate mentioned in Ramey (2021).

Figure 10: Subsample Stability of Elasticity Estimates

(a) Leave 4 quarters out (b) Leave 20 quarters out




*Notes*: SP-IV estimates of $\phi$ using the same specification as in row [1] of Table 1. Subsamples exclude 4 (left) or 20 (right) successive quarters starting with the quarter shown on the horizontal axis.

in all specifications. The estimates of $\phi_{ND}$, on the other hand, are very close to those in rows [2] and [3], and remain highly statistically significant. The only exception is in row [8]: Without the NASA appropriations, identification in the specification with both types of government R&D capital weakens to the point where the robust confidence intervals become very wide and include zero in all cases. This inference result is the only substantive difference between the weak-IV-robust inference methods and traditional Wald inference, which leads to the rejection of no spillovers even when excluding all NASA appropriations, see Appendix E.2. For the interested reader, Appendix E.3 provides the simultaneous confidence sets associated with the estimates in rows [6]-[9].

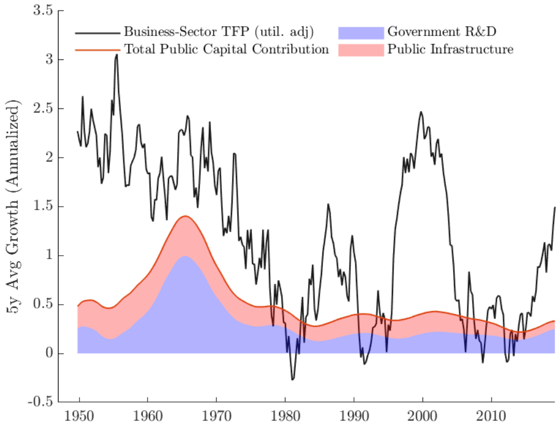
A key conclusion from Table 1 is that the various estimates of the production function elasticity to government R&D capital based on variation in nondefense R&D do not vary greatly, ranging from 0.09 to 0.12. Multiplying by the average postwar share of nondefense R&D capital of around 0.5, the estimates imply elasticities to nondefense R&D capital ranging from 0.045 to 0.06. The estimates of the nondefense elasticity are relatively precise (even under weak-instrument-robust inference) and highly statistically significant, with the exception of those in row [8]. Overall, the results point to sizable direct spillovers of nondefense government R&D on business-sector TFP. In contrast, the estimates based on variation in defense R&D vary considerably across specifications, from -0.13 to 0.20, and all come with wide confidence bands. Unlike for nondefense R&D, we cannot draw any sharp conclusions regarding the size—or even the sign—of any direct spillovers of defense R&D.

**Robustness** The estimates of the nondefense elasticities remain very similar after excluding all NASA appropriations, although they become less precise, see rows [2] and [8] in

Table 1. To further investigate the sensitivity of the results to outliers and particular policy episodes, Figure 10 plots SP-IV estimates of the nondefense R&D elasticity using the same specification as in row [1] of Table 1 after iteratively dropping parts of the sample. The left panel reports how the estimates evolve when leaving 4 successive quarters out of the sample on a rolling basis; the right panel shows the estimates when dropping five years from the sample. In both panels, the gray bands and blue lines show the 95 percent weak-instrument-robust and conventional Wald confidence bands, respectively. As the left panel shows, the estimation results do not change meaningfully in subsamples that omit 4 successive quarters. Unsurprisingly, there is more variability in the estimates across subsamples that omit five successive years from the sample. Nevertheless, the point estimates lie within a range of 0.06 to 0.015 and generally remain highly statistically significant. The exception is when the sample omits the late 1950s/early 1960s, corresponding to the escalation of the Cold War and the space race, in which case the statistical significance becomes more marginal. In a handful of subsamples, the robust inference procedures indicate that the estimates are no longer significant. Figure 10 illustrates that the large policy changes during that period are important for the precision of the estimates. At the same, no other historical episodes appear particularly important for the magnitude or precision of our baseline results.

The Online Appendix considers a number of additional robustness exercises that we summarize here. Appendix E.4 reports results for an alternative version of (8) that instead assumes constant elasticities to $\Delta k_t^{ND}$ and $\Delta k_t^D$. The results are broadly consistent with those in Table 1. After scaling appropriately for comparability, the estimates of $\phi_{ND}$ range from 0.06 to 0.13 but are generally somewhat smaller than those reported in Table 1. The estimates of $\phi_D$ for the alternative specification range from $-0.11$ to 0.12 and are all insignificant and imprecisely estimated. Appendix E.5 considers how the results change under different assumptions about the depreciation rates used to construct the capital stocks, which for R&D are difficult to measure in practice. The estimate $\hat{\phi}^{ND}$ ranges from 0.18, when all depreciation rates are assumed to be zero, to 0.09, when doubling the depreciation rates applied by the BEA. Assumptions about R&D depreciation are, therefore, quantitatively important for the elasticity estimates. Appendix E.6 considers the effect of possible mismeasurement of the elasticity of private R&D capital, which—as discussed earlier—is one potential source of upward bias. We find that $\hat{\phi}^{ND}$ varies little when making the business-sector R&D capital elasticity up to three times larger than the values in Fernald (2012), as the increase in private R&D spending following a nondefense shock is too small to meaningfully affect the estimates. Finally, Appendix E.7 considers specifications that also include private R&D capital, but these also yield little evidence of substantial bias due to possible mismeasurement of the elasticity of private R&D capital.

Figure 11: TFP Growth Contributions of Public Infrastructure and Government R&D




| | '47-'69 | '70-'89 | '90-'09 | '10-'21 |
|---|---|---|---|---|
| TFP growth | 1.98 | 0.98 | 1.15 | 0.87 |
| | | *a. Intermediate* $\eta$ | | |
| Infrastructure | 0.33 | 0.19 | 0.19 | 0.09 |
| R&D | 0.48 | 0.25 | 0.19 | 0.19 |
| | | *b. Low* $\eta$ | | |
| Infrastructure | 0.27 | 0.16 | 0.15 | 0.07 |
| R&D | 0.50 | 0.25 | 0.20 | 0.20 |
| | | *c. High* $\eta$ | | |
| Infrastructure | 0.50 | 0.29 | 0.28 | 0.14 |
| R&D | 0.44 | 0.22 | 0.18 | 0.18 |

*Notes*: The left panel shows the centered five-year moving average annualized growth rate of utilization-adjusted TFP from Fernald (2012) and the contributions of public capital assuming $\eta = 0.08$. The right panel tabulates averages across selected periods for different values of $\eta$.

## V. The Macroeconomic Returns to Government R&D

### A. Historical Contributions to TFP Growth

With the estimates of the TFP spillovers of government R&D in hand, we can assess the contribution of public capital accumulation to postwar business-sector TFP growth from a growth accounting perspective. When calculating the contributions of the different types of public capital, we assume that there are no TFP spillovers from defense R&D, i.e., $\phi_D = 0$. While the elasticity for defense R&D is imprecisely estimated, this assumption is consistent with the estimation results in Table 1. We also continue to assume that defense capital (i.e., defense equipment and structures) does not generate any TFP spillovers, as is the convention in the literature. The contribution of nondefense R&D is calculated as $\hat{\phi}_{ND} \times \left(s_{ND,t} \Delta k_t^{ND}\right)$. For $\hat{\phi}_{ND}$, we use the point estimates from row [1] in Table 1, which are in the middle of the range of estimates across the different specifications. The contribution of public infrastructure is calculated as $\eta \Delta q_t$. The figure in the left panel of Figure 11 shows the resulting contributions of government R&D and public infrastructure for $\eta = 0.08$. The right panel reports averages over selected time windows for each of the three values of $\eta$.

The main finding is that government R&D has contributed substantially to total TFP growth since WWII—accounting for roughly one-fifth of the total, on average—regardless of the value of $\eta$ within Ramey's (2021) plausible range. The contribution of government R&D is frequently similar to that of public infrastructure, and often larger. Between 1947 and 1969—when both government R&D and public infrastructure grew at a rapid pace—the combined contribution of growth in public capital accounts for 0.77 to 0.94 percentage points of average TFP growth of 1.98 percentage points. For the low value $\eta = 0.065$, the contribution of government R&D is almost twice as large as that of public infrastructure: 0.50 versus 0.27 percentage points, respectively. For the high value $\eta = 0.12$, government

R&D contributes only slightly less than public infrastructure: 0.44 percentage points vs. 0.50 percentage points, respectively. Relative to 1947-69, average TFP growth decelerated by 1.0 percentage points over 1979-89. The combined contribution of slower growth in public capital ranges from 0.36 to 0.43 percentage points as $\eta$ increases from low to high. The slowdown in government R&D alone explains between 0.22 (high $\eta$) and 0.25 (low $\eta$) percentage points, or one-fifth to one-quarter of the overall decline in TFP growth. According to our estimates, therefore, the retrenchment of government R&D in the 1970s and 1980s was at least as important for the slowdown in productivity as the slower pace of public infrastructure investment. For comparison, in a general equilibrium model calibrated with causal estimates of patent spillover elasticities, Dyèvre (2024) finds that declining government R&D spending accounts for roughly one-third of the deceleration in TFP growth over the postwar period. Our growth accounting results put this contribution at one-fifth to one-quarter (depending on $\eta$), so somewhat smaller than Dyèvre (2024).

The left panel in Figure 11 shows that government R&D spillovers were particularly important in the 1960s and early 1970s. A potential caveat to this finding is that the assumption of a constant $\eta$ throughout the entire postwar sample may not be realistic. Fernald (1999), for example, argues that road construction in the late 1950s and 1960s provided a one-time, unrepeatable, large productivity boost. If that is the case, our calculations likely overstate the contribution of government R&D relative to public infrastructure in that part of the sample. The left panel also shows that public investment—either in R&D or infrastructure—does not account for much of the high TFP growth immediately after WWII. An important caveat is that, by assumption, defense R&D plays no role in our decomposition. However, it is entirely plausible that the high levels of TFP growth in the 1950s are at least in part the result of wartime defense R&D spending, see also Antolin-Diaz and Surico (2024) or Gross and Sampat (2023).[15] Non-R&D defense procurement during WWII likely also had effects on productivity through learning by necessity, see Ilzetzki (2024), that may have persisted into the early postwar years. Finally, we note that the temporary pickup in TFP growth in the mid-1990s/early 2000s appears unrelated to public R&D or infrastructure capital, at least according to our accounting framework.

### B. Rates of Return to Government R&D

The production function elasticities reported in Table 1 can be translated into approximate rates of return to government R&D. The net rate of return on government R&D is $\rho_t^n = \rho_t - \delta_t$, where $\rho_t = \phi Y_t / K_t$ is the marginal product of $K_t$ (or gross return), $K_t$ is the government R&D capital stock, $Y_t$ is output, and $\delta_t$ is the depreciation rate of government R&D capital. We restrict attention to the return to nondefense R&D and use the estimates reported in the $\hat{\phi}/\hat{\phi}_{ND}$ columns of Table 1 for the calculations. To obtain an average gross

[15]Extending our methodology to WWII is difficult due to the lack of detailed defense spending estimates.

rate of return, we divide the elasticity estimates by the average ratio of government R&D capital to GDP (both in constant 2012 dollars), which is around 6 percent. We use real GDP rather than business-sector output for calculating the ratio based on the assumption that the productivity spillovers extend identically to production in the non-business sectors.

The rates of return calculated as just described are derived from the earlier estimates of the elasticity $\phi$, which is assumed to be constant over the estimation sample. A common alternative approach is instead to estimate $\rho$ as a constant, see e.g., Hall et al. (2010). Using $\Delta k_t \approx \Delta K_t/K_t$ and $\phi_t = \rho K_t/Y_t$, and substituting into (6) yields

$$\Delta \widetilde{tfp}_t \quad = \quad \rho \frac{\Delta K_t}{Y_t} + \Delta w_t \tag{9}$$

To estimate $\rho$, we follow the same methodology as in the previous section, but now with $\Delta K_t/Y_t$ as the endogenous regressor. Specifically, we estimate (9) using SP-IV regressions of the cumulative impulse responses of $\Delta\widetilde{tfp}$ and $\Delta K_t/Y_t$ to the nondefense R&D appropriations shocks. We again use real GDP rather than business-sector output as the measure of $Y_t$, which means that we assume that the spillovers are the same in the business and non-business sectors of the economy. We also consider specifications that explicitly allow for different returns on defense and nondefense government R&D capital:

$$\Delta \widetilde{tfp}_t \quad = \quad \rho_{ND} \frac{\Delta K_t^{ND}}{Y_t} + \rho_D \frac{\Delta K_t^{D}}{Y_t} + \Delta w_t \tag{10}$$

As before, we conduct inference using the weak-instrument-robust procedures of Kleibergen (2005), and only use forecast horizons at one-year intervals for the identifying moments to mitigate many-instrument problems.

Table 2 reports the estimates of the gross rate of return on nondefense R&D, both based on the elasticity estimates and those estimated directly by (9) or (10). The rows mirror the specifications in Table 1, with rows [1]-[3] reporting results for (9) and rows [4]-[7] reporting results for (10). Each row also reports the calculation of the rate of return based on the corresponding elasticity estimate in Table 1. Implied net returns can be obtained by subtracting $\delta \approx 0.16$, the average depreciation rate for government R&D calculated by the BEA.

As Table 2 shows, the implied rates of return to nondefense R&D are high. The reliable estimates range from around 140 percent to 210 percent depending on the specification, the assumed value of $\eta$, and the method of calculation. The SP-IV estimates of $\rho_{ND}$ are highly statistically significant regardless of the value of $\eta$. Similar to Table 1, the only exception is the specification with both types of government R&D and the narrative measure that excludes the NASA appropriations as the instrument, see row [6]. The point estimates in that specification are also much higher, but they are weakly identified and, as a result, unreliable. Lower values of $\eta$ imply that more of the TFP increase following a nondefense

TABLE 2: ESTIMATES OF THE RETURN TO GOVERNMENT R&D CAPITAL

| | Government R&D Measure | Instruments | Intermediate $\eta = 0.08$ | | Low $\eta = 0.065$ | | High $\eta = 0.12$ | |
|---|---|---|---|---|---|---|---|---|
| | | | $\hat{\phi}_{ND} \times \frac{Y}{K}$ | $\hat{\rho}_{ND}$ | $\hat{\phi}_{ND} \times \frac{Y}{K}$ | $\hat{\rho}_{ND}$ | $\hat{\phi}_{ND} \times \frac{Y}{K}$ | $\hat{\rho}_{ND}$ |
| [1] | Total | Exo ND | 1.85 | 1.71*** (1.07,2.22) | 1.91 | 1.77*** (1.13,2.26) | 1.67 | 1.57*** (0.91,2.11) |
| [2] | Total | Exo ND, No NASA | 1.94 | 1.60** (0.62,4.01) | 2.00 | 1.62** (0.69,4.03) | 1.77 | 1.53** (0.42,3.97) |
| [3] | Total | All ND | 1.79 | 1.58*** (1.04,2.08) | 1.86 | 1.63*** (1.10,2.12) | 1.62 | 1.44*** (0.88,1.98) |
| [4] | ND/D | Exo ND | 1.75 | 1.68** (0.23,3.20) | 1.81 | 1.74** (0.30,3.24) | 1.58 | 1.52** (0.08,3.11) |
| [5] | ND/D | Exo ND/D | 1.67 | 2.04** (0.12,3.79) | 1.73 | 2.10** (0.16,3.81) | 1.50 | 1.88** (0.01,3.70) |
| [6] | ND/D | Exo ND, No NASA | 1.92 | 6.84 (−2.00†,5.00†) | 1.98 | 6.91 (−2.00†,5.00†) | 1.75 | 6.65 (−2.00†,5.00†) |
| [7] | ND/D | All ND | 1.72 | 1.58** (0.27,2.90) | 1.78 | 1.64** (0.32,2.95) | 1.55 | 1.42** (0.11,2.81) |

*Notes:* Rows [1]-[3]: SP-IV estimates of $\rho$ in (9); rows [4]-[7]: SP-IV estimates of $\rho_{ND}$ in (10). 95 percent weak-instrument-robust intervals based on the KLM statistic of Kleibergen (2005) in parentheses. Test inversion is on a grid with endpoints −2 and 5, † denotes intervals constrained at these endpoints. Subvector inference in rows [6]-[9] uses the projection method. $*$, $**$ and $***$ denote statistical significance at 10, 5 and 1 percent levels, respectively. 'Exo ND/D': exogenous changes in nondefense/defense R&D appropriations. 'All ND/D': all changes in nondefense/defense R&D appropriations. 'No NASA': excluding all NASA appropriations. Sample: 1948Q1–2021Q4.

shock is attributed to R&D as opposed to public infrastructure, so the estimated returns are decreasing in $\eta$. However, the returns do not vary greatly across the plausible range for $\eta$ within each specification. Apart from the unreliable estimates in row [6], the estimated returns are also roughly the same regardless of whether they are derived from the elasticity estimates or estimated directly.

As discussed in the introduction, few other studies provide causal estimates of the returns to R&D using a methodology that is able to capture the various spillover margins at an aggregate level. Using changes in federal and state tax incentives for R&D for identification, Bloom et al. (2013) estimate a gross social rate of return on private R&D of 55 percent in a panel of U.S. firms. Jones and Summers (2022) use a stylized framework to calculate a gross social return to total R&D investment (private and public) of 67 percent based on aggregate U.S. data. Combining their estimated social return to total R&D with our baseline estimate of 171 percent in row [1] of Table 2, the private and public shares of R&D expenditures in the data, and an assumption that defense R&D has zero return, the implied social return on *private* R&D is 56 percent.[16] Thus, our estimates are consistent with the gap in social returns between the private R&D estimate in Bloom et al. (2013) and the total R&D estimate of Jones and Summers (2022). Other estimates of returns on private R&D in

[16]Based on average shares of total R&D expenditures for private R&D (0.55), federal defense R&D (0.24), and government nondefense R&D (0.21), the implied private social return is $(67\% - 0.21 \times 171\%)/(0.55) = 56.2\%$.

the literature vary widely and are more constrained in terms of their causal interpretation and the extent of spillover effects they are able to capture; see Hall et al. (2010) and Jones and Summers (2022) for further discussion.

Despite the methodological differences, our estimates of the return to government R&D of 140 to 210 percent appear meaningfully larger than the best available estimates for private R&D. One potential reason for the larger spillovers is the much stronger focus on fundamental research: For every dollar of federal nondefense R&D, on average, 34 cents goes to basic research. In contrast, for every dollar of privately funded R&D, just 6 cents is spent on basic research. As mentioned earlier, knowledge spillovers are also often deliberately promoted through the technology transfer programs and offices at the various federal agencies. In addition, Dyèvre (2024) documents that public R&D patents cite scientific publications much more heavily, are more likely to open new technological fields, are cited more widely across patent classes, and are disproportionally cited by smaller firms. All of these facts are suggestive of larger spillovers from publicly than privately funded R&D. Indeed, Dyèvre's (2024) patent elasticity estimates indicate that government R&D spillovers are two to three times as large as those of private R&D, which is roughly consistent with the difference between our return estimates and those of Bloom et al. (2013).

In terms of policy implications, our finding of large returns to government R&D implies substantial underinvestment of public funds in nondefense R&D. For comparison, the CBO estimates a gross return on public infrastructure capital of 12.4 percent and a net return of 9.2 percent after adjusting for depreciation (CBO 2021). Even after adjusting for the higher depreciation rates on R&D, the estimated returns in Table 2 substantially exceed those for public infrastructure, implying significant misallocation of public capital. Our estimates also suggest that federal investments in nondefense R&D are self-financing from the perspective of the federal budget, at least in the long run. Assuming a return of 171 percent, a \$1 long-run increase in government R&D capital would improve the budget as long as the additional tax revenue raised per dollar of additional GDP is at least 9 cents ($\delta/\rho = 0.16/1.71 = 0.09$), which is substantially below the historical ratio of federal tax revenues to GDP.

## VI. Avenues for Future Research

This paper contributes new causal evidence on the productivity effects of government funding for R&D by studying impulse responses to shocks to R&D appropriations for five major U.S. federal agencies. We use the impulse response estimates to structurally estimate the aggregate production function elasticity of government R&D capital. These estimates can be used to discipline quantitative models to study the long-run effects of public investment in research and innovation, as well as the optimal allocation of public capital between public infrastructure and knowledge capital. While we find evidence for direct productivity spillovers of nondefense R&D, the results for defense R&D are inconclusive, and it generally appears important to distinguish between investments in defense and nondefense research. Further distinctions between the various types of nondefense R&D funding, for instance, by type or agency, can be made to investigate the relative magnitude of the productivity spillovers. It is also possible to look at the effects of shocks to R&D appropriations in more disaggregated data and study the heterogeneous effects across firms or industries. The possible links between government R&D funding and overall trends in research productivity, as documented by Bloom et al. (2020), are also worth exploring. Another interesting avenue is to study R&D appropriations shocks as a potential deeper source of the 'technology news' shocks that are widely studied in macroeconomics, see also Jinnai (2014). Finally, our analysis has abstracted from global spillovers and possible international coordination of public investment in R&D. We leave these and other questions for future research.

## References


Akcigit, Ufuk, Douglas Hanley, and Nicolas Serrano-Velarde (2020). "Back to Basics: Basic Research Spillovers, Innovation Policy, and Growth". In: *The Review of Economic Studies* 88.1, pp. 1–43.

Alexopoulos, Michelle (2011). "Read All about It!! What Happens Following a Technology Shock?" In: *American Economic Review* 101.4, pp. 1144–79.

Antolin-Diaz, Juan and Paolo Surico (2024). *The Long-Run Effects of Government Spending*. Manuscript. London Business School.

Aschauer, David Alan (1989). "Is public expenditure productive?" In: *Journal of Monetary Economics* 23.2, pp. 177–200.

Azoulay, Pierre, Joshua S Graff Zivin, Danielle Li, and Bhaven N. Sampat (2019). "Public R&D Investments and Private-sector Patenting: Evidence from NIH Funding Rules". In: *The Review of Economic Studies* 86.1, pp. 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". In: *The Quarterly Journal of Economics* 138.2, pp. 895–954.

Beaudry, Paul and Franck Portier (2006). "Stock Prices, News, and Economic Fluctuations". In: *American Economic Review* 96.4, pp. 1293–1307.

Bianchi, Francesco, Howard Kung, and Gonzalo Morales (2019). "Growth, slowdowns, and recoveries". In: *Journal of Monetary Economics* 101, pp. 47–63.

Blanchard, Olivier and Roberto Perotti (2002). "An Empirical Characterization of the Dynamic Effects of Changes in Government Spending and Taxes on Output". In: *The Quarterly Journal of Economics* 117.4, pp. 1329–1368.

Bloom, Nicholas, Charles I. Jones, John Van Reenen, and Michael Webb (2020). "Are Ideas Getting Harder to Find?" In: *American Economic Review* 110.4, pp. 1104–1144.

Bloom, Nicholas, Mark Schankerman, and John Van Reenen (2013). "Identifying Technology Spillovers and Product Market Rivalry". In: *Econometrica* 81.4, pp. 1347–1393.

Bloom, Nicholas, John Van Reenen, and Heidi Williams (2019). "A Toolkit of Policies to Promote Innovation". In: *Journal of Economic Perspectives* 33.3, pp. 163–84.

Bom, Pedro R.D. and Jenny E. Ligthart (2014). "What Have We Learned from Three Decades of Research on the Productivity of Public Capital?" In: *Journal of Economic Surveys* 28.5, pp. 889–916.

Brunet, Gillian (2023). "When Does Government Spending Matter? Evidence from a New Measure of Spending". Manuscript.

Campbell, Sheila and Chad Shirley (2018). *Estimating the Long-Term Effects of Federal R&D Spending: CBO's Current Approach and Research Needs*. Tech. rep. Congressional Budget Office. URL: www.cbo.gov/publication/54089.

Cascaldi-Garcia, Danilo and Marija Vukotić (2022). "Patent-Based News Shocks". In: *The Review of Economics and Statistics* 104.1, pp. 51–66.

CBO (2016). *The Macroeconomic and Budgetary Effects of Federal Investment*. Tech. rep. Congressional Budget Office. URL: www.cbo.gov/publication/51628.

— (2021). *Effects of Physical Infrastructure Spending on the Economy and the Budget Under Two Illustrative Scenarios*. Tech. rep. Congressional Budget Office. URL: www.cbo.gov/publication/57327.

Cloyne, James, Joseba Martinez, Haroon Mumtaz, and Paolo Surico (2024). *Corporate Taxes, Innovation, and Productivity*. manuscript. London Business School.

Comin, Diego and Mark Gertler (2006). "Medium-Term Business Cycles". In: *American Economic Review* 96.3, pp. 523–551.

De Lipsis, Vincenzo, Matteo Deleidi, Mariana Mazzucato, and Paolo Agnolucci (2022). *Macroeconomic Effects of Public R&D*. Working Paper 2023-02. UCL Institute for Innovation and Public Purpose.

Dyèvre, Arnaud (2024). *Public R&D Spillovers and Productivity Growth*. Manuscript. London School of Economics.

Elmendorf, Douglas, Glenn Hubbard, and Heidi Williams (2024). "Dynamic Scoring: A Progress Report on Why, When, and How". In: *Brookings Papers on Economic Activity*.

Fama, Eugene F. (1990). "Stock Returns, Expected Returns, and Real Activity". In: *The Journal of Finance* 45.4, pp. 1089–1108.

Fernald, John G. (1999). "Roads to Prosperity? Assessing the Link between Public Capital and Productivity". In: *American Economic Review* 89.3, pp. 619–638.

— (2012). *A Quarterly, Utilization-Adjusted Series on Total Factor Productivity*. Working Paper Series 2012-19. Federal Reserve Bank of San Francisco.

Fieldhouse, Andrew J. and Karel Mertens (2023). *A Narrative Analysis of Federal Appropriations for Research and Development*. Dallas Fed Working Paper 2316.

Fieldhouse, Andrew J, Karel Mertens, and Morten O. Ravn (2018). "The Macroeconomic Effects of Government Asset Purchases: Evidence from Postwar U.S. Housing Credit Policy". In: *The Quarterly Journal of Economics* 133.3, pp. 1503–1560.

Galí, Jordi (1999). "Technology, Employment, and the Business Cycle: Do Technology Shocks Explain Aggregate Fluctuations?" In: *American Economic Review* 89.1, pp. 249–271.

Gross, Daniel P. and Bhaven N. Sampat (2023). "America, Jump-Started: World War II R&D and the Takeoff of the US Innovation System". In: *American Economic Review* 113.12, pp. 3323–56.

Gruber, Jonathan and Simon Johnson (2019). *Jump-Starting America: How Breakthrough Science Can Revive Economic Growth and the American Dream*. New York, NY: PublicAffairs, Hachette Book Group.

Gullo, Theresa, Benjamin Page, David Weiner, and Heidi L. Williams (2024). "Estimating the Economic and Budgetary Effects of Research Investments". In: *Tax Policy and the Economy, Volume 39*. Ed. by Robert A. Moffitt. University of Chicago press.

Hall, Bronwyn H., Jacques Mairesse, and Pierre Mohnen (2010). "Chapter 24 - Measuring the Returns to R&D". In: *Handbook of the Economics of Innovation, Volume 2*. Ed. by Bronwyn H. Hall and Nathan Rosenberg. Vol. 2. Handbook of the Economics of Innovation. North-Holland, pp. 1033–1082.

Ilzetzki, Ethan (2024). "Learning by Necessity: Government Demand, Capacity Constraints, and Productivity Growth". In: *American Economic Review* 114.8, 2436–71.

Jinnai, Ryo (2014). "R&D Shocks and News Shocks". In: *Journal of Money, Credit and Banking* 46.7, pp. 1457–1478.

Jones, Benjamin F. and Lawrence H. Summers (2022). "A Calculation of the Social Returns to Innovation". In: ed. by Austan Goolsbee and Benjamin Jones. NBER Innovation and Public Policy, 13–59.

Jordà, Òscar (2005). "Estimation and Inference of Impulse Responses by Local Projections". In: *American Economic Review* 95.1, pp. 161–182.

Kantor, Shawn and Alexander T. Whalley (2024). *Moonshot: Public R&D and Growth.* NBER Working Paper 31471.

Kelly, Bryan, Dimitris Papanikolaou, Amit Seru, and Matt Taddy (2021). "Measuring Technological Innovation over the Long Run". In: *American Economic Review: Insights* 3.3, pp. 303–20.

Kleibergen, Frank (2005). "Testing Parameters in GMM Without Assuming that They Are Identified". In: *Econometrica* 73.4, pp. 1103–1123.

Kogan, Leonid, Dimitris Papanikolaou, Amit Seru, and Noah Stoffman (2017). "Technological Innovation, Resource Allocation, and Growth". In: *The Quarterly Journal of Economics* 132.2, pp. 665–712.

Lazarus, Eben, Daniel J. Lewis, James H. Stock, and Mark W. Watson (2018). "HAR Inference: Recommendations for Practice". In: *Journal of Business & Economic Statistics* 36.4, pp. 541–559.

Leeper, Eric M., Todd B. Walker, and Shu-Chun Susan Yang (2013). "Fiscal Foresight and Information Flows". In: *Econometrica* 81.3, pp. 1115–1145.

Lewis, Daniel J. and Karel Mertens (2023). *Dynamic Identification Using System Projections and Instrumental Variables.* Working Paper 2204. Federal Reserve Bank of Dallas.

Mazzucato, Mariana (2018). *The Entrepreneurial State: Debunking Public vs. Private Sector Myths.* New York, NY: PublicAffairs, Hachette Book Group.

Mertens, Karel and Morten O. Ravn (2013). "The Dynamic Effects of Personal and Corporate Income Tax Changes in the United States". In: *American Economic Review* 103.4, pp. 1212–47.

Miranda-Agrippino, Silvia, Sinem Hacioğlu Hoke, and Kristina Bluwstein (2022). *Patents, News, and Business Cycles.* Discussion Paper 15062. Centre for Economic Policy Research.

Moretti, Enrico, Claudia Steinwender, and John Van Reenen (2021). *The Intellectual Spoils of War? Defense R&D, Productivity and International Spillovers.* Working Paper 26483. National Bureau of Economic Research.

Moris, Francisco and Christopher Pece (2022). *Definitions of Research and Development: An Annotated Compilation of Official Sources.* Tech. rep. NCSES 22-209. National Center for Science and Engineering Statistics.

Munnell Alicia, H. (1990). “Why has Productivity Growth Declined? Productivity and Public Investment”. In: *New England Economic Review*, pp. 3–22.
Myers, Kyle R. and Lauren Lanahan (2022). “Estimating Spillovers from Publicly Funded R&D: Evidence from the US Department of Energy”. In: *American Economic Review* 112.7, pp. 2393–2423.
Newey, Whitney K. and Frank Windmeijer (2009). “Generalized Method of Moments with Many Weak Moment Conditions”. In: *Econometrica* 77.3, pp. 687–719.
Ramey, Valerie A. (2011). “Identifying Government Spending Shocks: It’s all in the Timing”. In: *The Quarterly Journal of Economics* 126.1, pp. 1–50.
— (2021). “The Macroeconomic Consequences of Infrastructure Investment”. In: *Economic Analysis and Infrastructure Investment*. Ed. by Edward L. Glaeser and James M. Poterba. Chicago: University of Chicago Press.
Ramey, Valerie A. and Sarah Zubairy (2018). “Government Spending Multipliers in Good Times and in Bad: Evidence from US Historical Data”. In: *Journal of Political Economy* 126.2, pp. 850–901.
Romer, Christina D. and David H. Romer (1989). “Does Monetary Policy Matter? A New Test in the Spirit of Friedman and Schwartz”. In: *NBER Macroeconomics Annual* 4, pp. 121–170.
— (2010). “The Macroeconomic Effects of Tax Changes: Estimates Based on a New Measure of Fiscal Shocks”. In: *American Economic Review* 100.3, pp. 763–801.