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

Andrew J. Fieldhouse
Mays Business School
Texas A&M University

Karel Mertens†
Federal Reserve Bank of Dallas
CEPR

First version: May 5, 2023
This version: December 14, 2023

Abstract

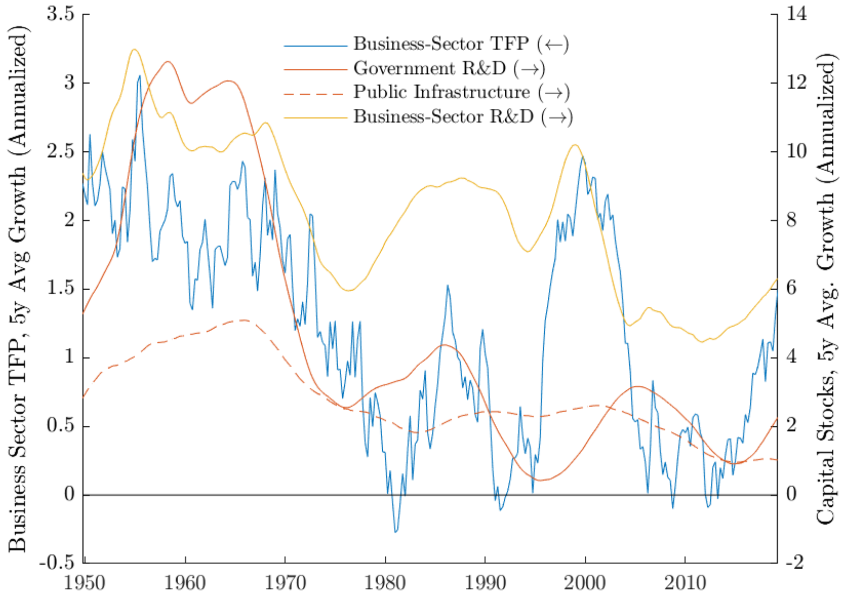
Based on a narrative classification of all significant postwar changes in R&D appropriations for five major federal agencies, we find that an increase in nondefense R&D appropriations leads to increases in various measures of innovative activity and higher business-sector productivity in the long run. 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 150 to 300 percent over the postwar period. The estimates indicate that government-funded R&D accounts for one quarter 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 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, the Barcelona Summer Forum, the Central Bank of Ireland, Clark University, the Federal Reserve Banks of Atlanta and Dallas, Ghent University, Leuven University, the Liberal Arts Macroeconomics Conference, the Midwest Macro Conference, the Lundquist College of Business, the Mays Business School, Middlebury College, Oxford University, and Texas A&M University for helpful feedback. We thank Roni Lezama, Annie Midthun, Tai Nguyen, Michaela Sullivan, and Nicholas Wagg for their excellent research assistance. Andrew Fieldhouse gratefully acknowledges funding received from Middlebury College for this project. 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.

†*Contact*: 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 the discussion in Gruber and Johnson (2019) or 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 advancements in military know-how are unlikely to be disseminated as quickly in order to maintain military superiority. For this reason, we distinguish throughout the analysis between defense and nondefense R&D. We find that a positive shock to appropriations for nondefense R&D robustly leads to a delayed and gradual increase in aggregate 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 increases in the level of TFP 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 larger spillovers. We find that nondefense shocks lead to relatively larger increases in funding for more fundamental 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.

We find that positive shocks to R&D appropriations for both defense and nondefense activities lead to higher private investment in R&D. As in the theoretical framework of Akcigit et al. (2020), this suggests that private and public R&D investments indeed act as complements rather than substitutes. However, the increases in private R&D are relatively small, particularly in response to nondefense shocks. 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 of 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. We consider values of this elasticity between 0.065 and 0.12, the range deemed plausible by Ramey (2021) in a recent review of the existing evidence. 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 responses to nondefense R&D shocks, the point estimates of the production function elasticity to total government R&D capital across various specifications lie within a relatively tight range of a value of 0.12, 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.

Our estimates of the production function elasticities imply that nondefense government R&D accounts on average for about one-quarter 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 0.38 to 0.45 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.

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 between 150 and 300 percent 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,

[2] The value of 0.12 for the elasticity to total government R&D capital 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.

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 recent empirical studies focus on industry-specific spillovers or patent responses to specific government R&D programs. For instance, Azoulay et al. (2018) find that NIH spending spurs the generation of private patents; Myers and Lanahan (2022) find large private R&D spillovers from the DOE's Small Business Innovation Research program; Kantor and Whalley (2022) find persistent manufacturing output and productivity 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; and Moretti et al. (2019) find positive spillovers from defense R&D to private R&D and productivity growth in a panel study of defense R&D spending across OECD countries. Each of these studies provides evidence for some of the spillovers that we aim to measure collectively.

Our paper is also related to several recent studies of the longer-run macroeconomic effects of fiscal policy shocks. Cloyne et al. (2022), 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 (2022) 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 are associated with shocks that expand the share of government spending going to R&D, which they identify by maximizing the variance of government R&D spending at forecast horizons of up to one year. 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 (2022) 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 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

[3]For example, Bloom et al. (2013) use firm-level accounting data and changes in R&D tax incentives to identify a 55 percent social rate of return to R&D. See Hall et al. (2010) or Jones and Summers (2020) for overviews of the evidence.

function elasticity of government R&D capital that can be used to separately study the role of intangible public capital in quantitative growth models. These estimates are also useful for budgetary analyses of fiscal policy initiatives (e.g., CBO 2016; CBO 2021).

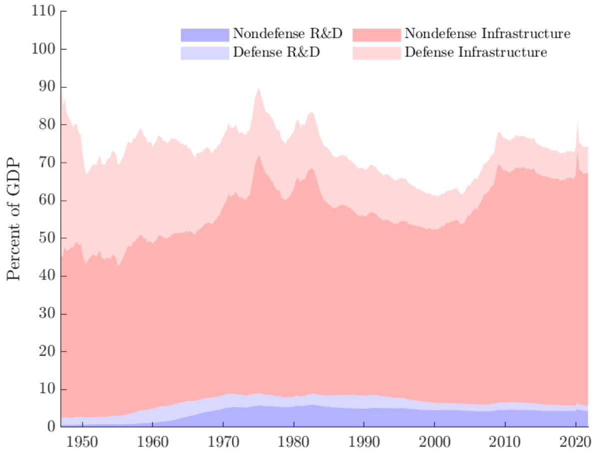
I. Measurement, Definitions, and Facts

The measures of public capital used in this paper are based on data from the Bureau of Economic Analysis (BEA). Specifically, we use data on gross investment from the National Income and Product Accounts (NIPAs) to construct quarterly series of the value of government fixed assets (at real cost) that are consistent with the annual series in the BEA's Fixed Asset Account, see Appendix A for details. 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. 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 NIPAs; we use the term 'R&D capital' as such throughout the rest of the paper.[4] We refer to the aggregate of (iii) and (iv) as 'government R&D capital'. R&D expenditures are measured in the NIPA 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.

The expenditure data underlying the BEA measures of R&D capital are constructed primarily 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 NSF defines R&D as the "creative and systematic work undertaken in order to increase the stock of knowledge ... and to devise new applications of available knowledge." This wide umbrella for spending on innovative activity is typically separated into three types: basic research, applied research, and experimental development work. The NSF defines basic research as experimental or theoretical work pursuing knowledge "without specific applications toward processes or products" whereas experimental development work is defined as "systematic work, drawing on knowledge gained from research and practical experience and producing additional knowledge, which is directed to producing new products or processes or to

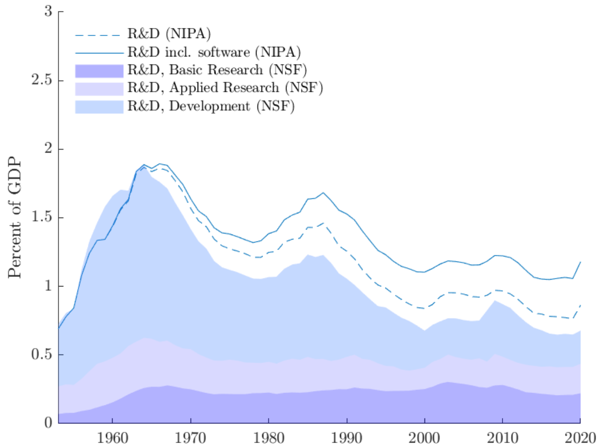
[4]In Appendix D.5, we also consider alternative measures of R&D capital based on alternative values for the depreciation rates than those used by the BEA.

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).

improving existing products or processes." Falling between these two, applied research is defined as "original investigation undertaken in order to acquire new knowledge... directed primarily towards a specific practical aim or objective" (NSF 2022).

Figure 3 plots the NCSES measures of government R&D spending by type, along with NIPA totals for comparison. Government spending on basic and applied research each averaged 0.23 percent of GDP over the sample period shown, while experimental development averaged 0.61 percent. Government 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, government R&D expenditures in the NSF surveys do not align perfectly with the corresponding series in the national accounts, as the BEA adjusts the NSF 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] Note that not all spending labeled research or development in other data sources, such as the appropriations data we use in our analysis, 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.

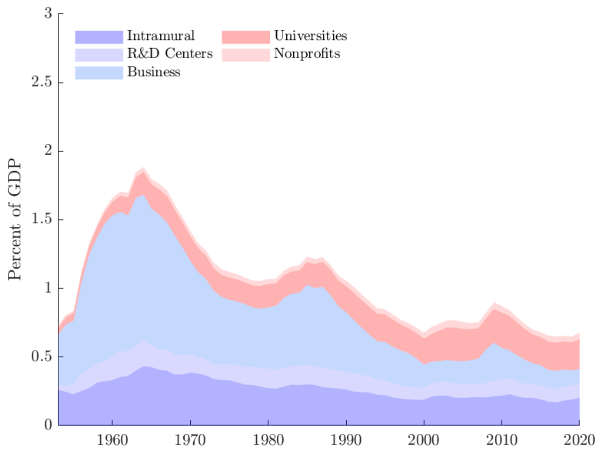
Figure 4 plots NCSES data on government R&D spending broken out by performing sectors and the major funding agencies. Panel (a) shows that the bulk of government R&D

[5] Over the same period, average private expenditures are 0.14 percent of GDP on basic research, 0.30 percent of GDP on applied research, and 0.97 percent of GDP on development (Source: NCSES, National Patterns of R&D Resources).

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

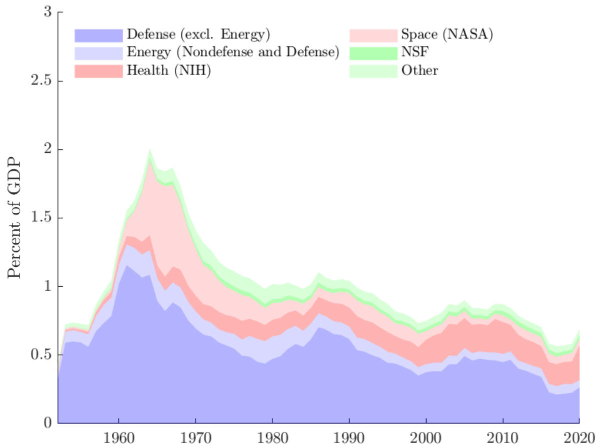
Figure 4: Government R&D Expenditures

(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).

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. During the height of the Cold War, most government-funded R&D was performed by businesses, but the share has fallen substantially since, and a steadily growing share is performed at universities. Government funds for R&D are provided largely by the federal government—more than 90 percent on average in the postwar period; the remainder consists mainly of funding by state and local governments 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 can be attributed to Congress reversing course on funding for these agencies after the nuclear triad was deployed and the moon landing was successfully completed. 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 (e.g., civilian energy and physics research); in the NIPAs, DOE's national security functions are included in defense R&D. In recent decades, NIH funding for medical research has gradually grown in importance. The final agency engaged in significant R&D funding is the NSF. Various other federal agencies also provide funding for R&D 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 on aggregate productivity is based on novel empirical measures of exogenous variation in federal funding for R&D. As is well known in the literature, an important identification concern is that changes in fiscal policy are often anticipated, and mistiming the arrival of news about fiscal policy can lead to misleading results (Ramey 2011; Mertens and Ravn 2013; Leeper et al. 2013). To address these concerns, we rely on time series of all enacted appropriations authorizing future federal R&D expenditures, and not just on current R&D expenditures as in Antolin-Diaz and Surico (2022) or De Lipsis et al. (2022).[7]

The other identification concern is that policy changes reflect systematic reactions by policymakers to macroeconomic developments that independently affect innovative activity and aggregate productivity growth. 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.[8] Second, to guard against the possibility that R&D policy responds systematically to other longer-term drivers of productivity trends, we embed the narrative measures in empirical models that remove predictable variation in future productivity growth through a wide variety of lagged controls at a quarterly frequency. As we will show, neither the narrative identification step nor the choice of controls will prove crucial for our main empirical finding that nondefense government R&D raises TFP in the long run, which likely helps explain why our results broadly agree with those of Antolin-Diaz and Surico (2022) or De Lipsis et al. (2022) despite the various methodological differences. Before we describe the econometric methodology in full detail, the rest of this section first discusses the dataset on appropriations as well as 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 restrict attention to congressional appropriations for R&D activities. To construct a time series of federal R&D appropriations, we rely on information in the *Budget of the U.S. Government* and its appendices. Specifically, we collect information on all enacted appropriations for the budgetary accounts 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 federal agencies discussed in Figure 4: DOD, DOE, NASA, NIH, and NSF.[9] Together, these five agencies typically account for roughly 87 to 93 percent of total federal R&D spending in any given fiscal year. For each agency, we obtain the appropriations from

[7]See also Brunet (2022) on the estimation of fiscal multipliers using budget authority rather than outlays.

[8]Examples of similar empirical approaches include applications to monetary policy (Romer and Romer 1989; Romer and Romer 2023), government spending (Ramey 2011; Ramey and Zubairy 2018), federal tax policies (Romer and Romer 2010; Cloyne 2013), and housing credit policies (Fieldhouse et al. 2018).

[9]The Atomic Energy Commission and Energy Research and Development Administration are included as precursors to the Department of Energy.

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 data we collect reflects all enacted annual appropriations bills adjusted for any supplemental appropriations, subsequent transfers between accounts, or sequestration cuts. We date the appropriations to the quarter they take effect, either the start of that fiscal year or when the appropriations bill was subsequently enacted. As such, most changes in appropriations are dated to the first quarter of the fiscal year. To match defense and nondefense spending in the NIPAs, we separate the appropriations for DOD and for the national security functions of DOE from all other nondefense appropriations. References to all data sources by agency/year are available in Fieldhouse and Mertens (2023).

B. Narrative Classification

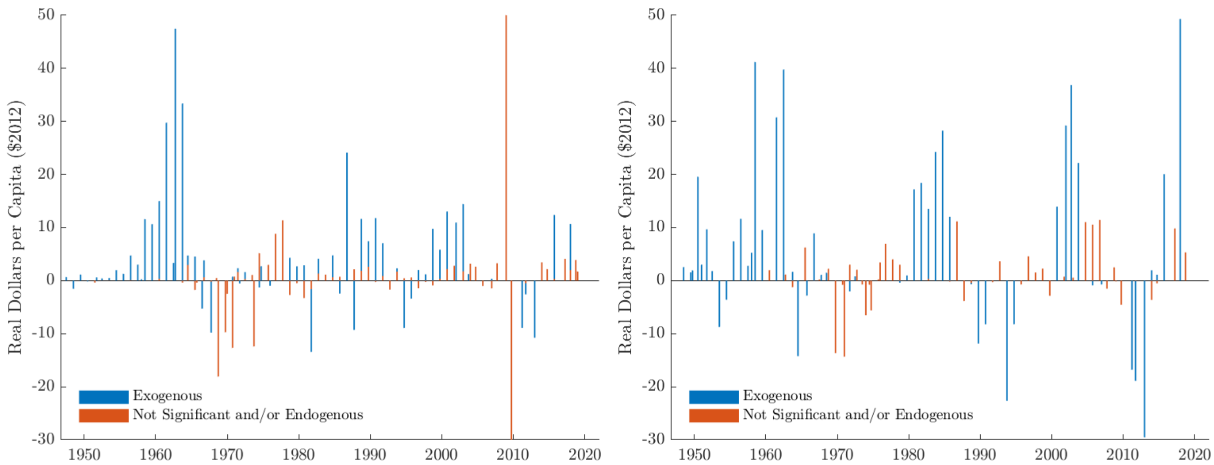
One potential reason for endogeneity problems is that changes in R&D appropriations may be correlated with business cycle shocks. Comin and Gertler (2006) and Bianchi et al. (2019), for example, argue that expansionary business cycle shocks can raise aggregate productivity at longer horizons through endogenous growth channels, while Ilzetzki (2022) provides evidence that high capacity utilization during WWII spurred innovation out of necessity. Government R&D appropriations may be procyclical given that there is more room in government budgets during booms. On the other hand, R&D spending may also rise in recessions if increases in appropriations for R&D are systematically folded into larger fiscal stimulus packages.

As we will rely on quarterly regression specifications that include lagged cyclical indicators as controls, one could appeal to lags in the policymaking process for identification, as in for example Blanchard and Perotti (2002). However, as including lagged cyclical indicators as controls may not suffice to remove all sources of cyclical endogeneity, we prefer to conduct our analysis with a subset of changes in appropriations that are classified as exogenous by a narrative analysis. More specifically, for each of the five agencies, we conduct a narrative analysis for all fiscal years 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 for two reasons. First, it is easier to infer legislative intent from available historical sources for the more meaningful deviations from current policy. Second, the focus on larger changes substantially reduces the number of agency-fiscal year pairs to analyze. In total, we classify 218 agency-fiscal year pairs with significant real changes in appropriations over the FY1947-2019 sample. Roughly one-third of the policy changes involve decreases in real appropriations for R&D and two-thirds are increases.

For each of the 218 significant agency-fiscal year changes, we rely on a variety of primary and secondary sources to understand the context and motivation. Specifically, we study the *Budget of the United States Government*, the *State of the Union address*, and

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 appropriations 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*, see Fieldhouse and Mertens (2023).

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. Finally, we also scan *CQ Almanac* and 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 the various sources, we classify every significant change in real R&D appropriations for each agency as either 'exogenous' or 'endogenous'. Endogenous policy changes are those that are primarily motivated by short-run economic considerations. Examples are increases in R&D spending that are part of a broader fiscal stimulus package (e.g., as in the American Recovery and Reinvestment Act of 2009), increases in energy R&D in response to oil shocks (e.g., creating the Energy Research and Development Administration in 1974), or cuts to R&D spending as part of broader austerity measures intended to curb short-run inflationary pressures. Exogenous policy changes are instead motivated by a variety of other considerations without clear macroeconomic relevance in the short run. For nondefense R&D, examples include policymakers' general concerns about the adequacy of the science, technology, and engineering base (e.g., creating 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 creating NASA), or initiatives with mixed diplomatic/scientific objectives (e.g., International Space Station). For defense R&D, examples include concerns about the adequacy of strategic capabilities relative to geopolitical rivals (e.g., Sputnik crisis), policy preferences of a new administration (e.g., Reagan's military buildup), evolving national security threats (e.g., Global War on Terror), or ratifying or exiting non-proliferation treaties (e.g., exiting the Anti-Ballistic Missile Treaty). Long-term deficit reduction packages often cut nondefense and/or defense R&D appropriations (e.g., Budget Control Act of 2011). We classify such policies as exogenous if the intent is long-term fiscal sustainability rather than curbing near-term inflationary pressures.

Figure 5 shows the time series of the changes in defense and nondefense appropriations, expressed in 2012 dollars per capita for ease of comparison across the sample period. The blue bars show those changes that are classified as exogenous in the narrative analysis, aggregated over all five agencies. Appendix B presents the same figures for each agency separately. For the interested reader, Fieldhouse and Mertens (2023) provides an overview of postwar federal R&D policy along with details and data sources for each policy change.

C. Orthogonalized Narrative Measures for Changes in Defense and Nondefense Appropriations

The knowledge spillovers from defense and nondefense R&D are potentially very different, if only because defense R&D usually aims to maintain America's military advantage and is often classified. One slight complication to isolating their separate effects empirically is that the changes in defense and nondefense R&D appropriations shown in Figure 5 are positively correlated, i.e., an increase in appropriations for one category tends to be accompanied by an increase in the other. Specifically, the correlation between all changes in defense and nondefense appropriations is 0.31, and the correlation across the exogenous policy changes is also 0.31. To better understand any underlying differences, it is useful to estimate the causal effects of more idiosyncratic movements in each category of government R&D. To that end, we construct versions of the narrative measures that are orthogonalized with respect to one another. More specifically, let $\Delta a_t^{exo,i}$ denote the narrative measures of exogenous changes in appropriations for $i = D, ND$ (defense/nondefense) in quarter t, as shown in blue in Figure 5. The orthogonalized narrative measures are the residual z_t^i in the following regression,

$$\text{(1)} \qquad \frac{\Delta a_t^{exo,i}}{K_{t-4}^i} = a_i + b_i \frac{\Delta a_t^{exo,-i}}{K_{t-4}^{-i}} + z_t^i \quad , \quad i = D, ND \ .$$

To construct the orthogonalized narrative measures, we express the (constant) dollar changes in appropriations in category i as a fraction of the total real value of the government R&D capital stock in that budget category four quarters earlier, K_{t-4}^i. We scale the changes in R&D appropriations by the real capital stocks as we are interested in elasticities to government R&D capital. To avoid introducing any sources of endogeneity, we scale by the one-year lagged capital stocks, although this matters very little for the results. By construction, the sample correlation between z_t^i and the exogenous appropriation change in the other category, $\Delta a_t^{exo,-i}/K_{t-4}^{-i}$, is zero. The orthogonalized narrative measure z_t^i, therefore, represents an exogenous innovation in government R&D appropriations for category i at time t, but leaving appropriations in the other category $-i$ contemporaneously unchanged. The impulse responses identified with the orthogonalized narrative measures will have the interpretation as the impact of a change in R&D funding targeting one category while leaving appropriations for the other category unchanged on impact (but not necessarily in future quarters). In practice, the orthogonalization step turns out to matter very little for the results, see Appendix C.2. Note that the positive correlation between defense and nondefense measures implies that the two measures potentially both contain useful identifying variation in defense and nondefense R&D capital. Our estimation of the production function elasticities and rates of return will therefore also include specifications that simultaneously use both narrative measures (without the orthogonalization) for identification.

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 associated with unanticipated changes—or 'shocks'—to defense and nondefense R&D appropriations. 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 use Jordà (2005) local projections to estimate responses at forecast horizons $h = 0, ..., H-1$ of up to 15 years ($H = 60$ quarters).[10] The impulse response for an outcome variable y_t at horizon h estimated by local projections is simply the OLS coefficient in a direct forecasting regression of y_{t+h} on the period t value of the orthogonalized narrative measures, z_t^i. This estimation approach makes no ex-ante assumptions regarding the lags between R&D spending and the impact on productivity. Because changes in R&D appropriations are serially correlated, as seen in Figure 5, we include information about past R&D appropriations in the regression. Specifically, we include $p = 4$ quarterly lags of $\ln(a_t^i)$, where a_t^i is the cumulative sum of all past (constant dollar) changes in 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 Appendix C.4 shows that additionally including lags of z_t^i has little effect on the results. We also include $p = 4$ lags of the outcome variable y_t in all specifications. Unless mentioned otherwise, the estimation sample consists of 74 years of quarterly observations from 1948Q1 through 2021Q4.

[10]Vector autoregressions (VARs) are a common alternative for impulse response estimation. As shown in Plagborg-Møller and Wolf (2021), local projections avoid potential misspecification in finite-order VAR-based impulse response estimators at forecast horizons beyond the VAR lag length. Appendix C.5 shows that VAR-based estimates to the appropriations shocks are nevertheless similar to the LP estimates.

In practice, we estimate the following local projections for $h = 0, ..., H-1$ using OLS:

$$(2) \quad \sum_{j=0}^{3} \left(\frac{1}{4} \times y_{t+h-j} \right) = c_h + \gamma_h z_t^i + \sum_{j=1}^{p} \beta_h^j \ln a_{t-j}^i + \sum_{j=1}^{p} \delta_h^j y_{t-j} + \sum_{j=1}^{p} \zeta_h^{j\prime} x_{t-j} + v_{t+h}$$

where $p = 4$, y_{t+h} is the outcome variable of interest at horizon h (e.g., utilization-adjusted TFP), v_{t+h} is a residual at forecast horizon h, and the sequence $\{\gamma_h\}_{h=0}^{H-1}$ contains the impulse response coefficients.

Two features of (2) warrant further discussion: First, the left-hand side is a four-quarter backward moving average, $\sum_{j=0}^{3} \left(\frac{1}{4} \times y_{t+h-j}\right)$, rather than just the quarterly observation y_{t+h}. The averaging smooths out some of the quarterly noise in the impulse response estimates, but is otherwise not important: the estimation 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.

Second, the specification in (2) allows for the inclusion of lags of additional control variables, x_t. As is well known, including lagged predictors of the outcome variables as controls in local projections can serve multiple purposes. One is that, even when identification is valid without conditioning on lagged predictors of y_{t+h}, including these predictors generally sharpens inference on the impulse response estimates by reducing the variance of the forecast residuals, v_{t+h}. Another is that adding a suitable set of lagged controls helps to eliminate past influences on the outcome variable that may be correlated with the regressor of interest, and otherwise would lead to endogeneity bias.

As discussed earlier, one of the controls included in x_t is a cyclical indicator to eliminate any remaining cyclical sources of endogeneity in the narrative measures. In the baseline set of controls x_t, we include the capacity utilization rate from the Fernald (2012) dataset, which captures variation in both labor effort and the workweek of capital and is strongly correlated with other coincident cyclical indicators. Adding the unemployment rate or the output gap has very little impact on the results, see Appendix C.3.

Even if the narrative classification and cyclical controls successfully address the short-run sources of endogeneity that are typically of greatest concern in the identification of fiscal shocks, it is not clear that they are adequate to address potential longer-run sources of policy endogeneity. R&D policy may, for example, respond to productivity, demographic, or other secular trends. A related possibility is that policy responds to the arrival of new ideas and nascent technologies that, even in the absence of 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 also include real government and business-sector R&D capital (both in log-levels) in x_t. Including R&D capital stocks, 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. 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. The final element in the baseline control set x_t is the defense spending news variable of Ramey and Zubairy (2018). We include news about total defense spending to remove additional predictable variation in defense R&D, and potentially also in nondefense R&D arising through complementarities or government budget constraints.

In Appendix C.3, we establish robustness to numerous additions to this baseline set of controls, including a variety of additional 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).

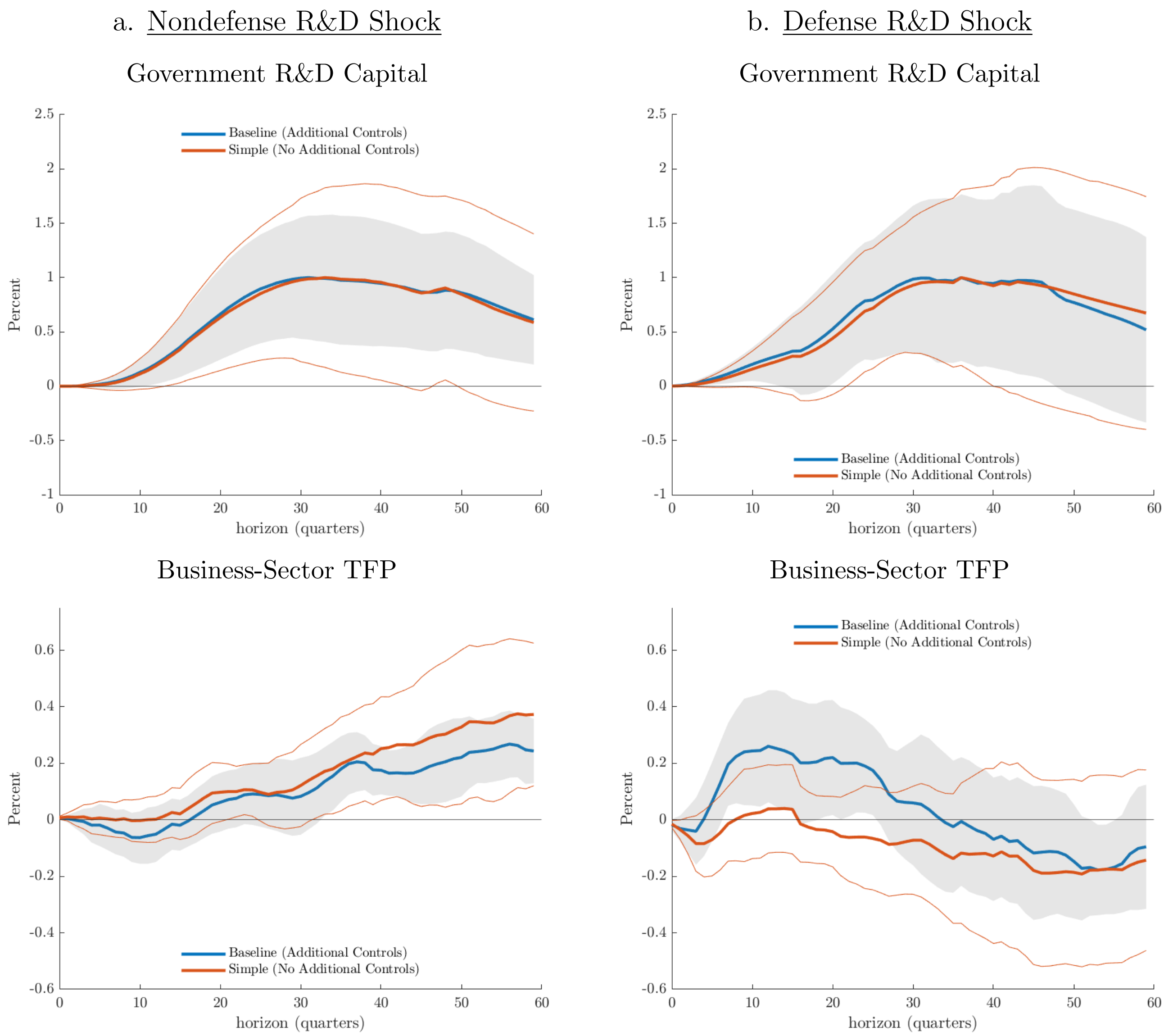
B. Government R&D and TFP After Shocks to R&D Appropriations

Figure 6 presents 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 (2). Each panel shows results for the baseline specification, i.e., with the five additional controls in x_t described above. To assess the importance of including these additional controls, the panels also report results from a simpler specification without the lags of any of the variables in x_t. For ease of comparison with the production function elasticities presented later, the responses are scaled to imply a one percent peak increase in total government R&D capital. Inference is based on the heteroskedasticity and autocorrelation-robust (HAR) confidence bands recommended by Lazarus et al. (2018).[11]

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 very 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

[11] Appendix C.6 shows that our main result remains unchanged when using other inference procedures (specifically Ecker-Huber-White or the Montiel Olea and Plagborg-Møller (2021) wild bootstrap).

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



Notes: Estimates based on (2) using the orthogonalized narrative measure of changes in federal nondefense (left panel) and defense (right panel) R&D appropriations, see (1). 'Baseline' includes additional lagged controls described in the main text. Lazarus et al. (2018) HAR bands are at the 5 percent significance level. Impulses scaled to imply a 1 percent peak increase in government R&D capital. Sample: 1948Q1–2021Q4.

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 increase in government R&D capital after a nondefense shock is highly statistically significant for all horizons except shorter ones, which indicates that the narrative nondefense measure is a strong predictor of future government R&D spending. The response of government R&D capital to a defense shock is also significant at the 5 percent level at horizons between 5 and 11 years, but the confidence bands are wider than for the nondefense shock. The fact that congressional appropriations are strongly predictive for future government R&D spending implies that the spending changes are potentially anticipated well in advance. Basing identification on variation in appropriations rather than spending is, therefore, preferable to avoid possible bias due to anticipation effects. For both shocks, the point estimates vary little across the specifications with and without the additional controls. The main effect of the additional controls is to substantially sharpen inference for the government R&D capital response to a nondefense shock.

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 appropriations leads to a gradual increase in business-sector TFP. Moreover, the TFP increase becomes highly statistically significant at longer horizons. In our 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 simpler specification, the TFP response is somewhat larger, up to around 0.35 percent at the end of the forecast horizon. Including the additional controls again considerably increases the precision of the estimates, but the TFP response is overall similar in shape, and it is 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 is meaningfully different 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 the baseline specification, the longer-run decrease in TFP is significant at the 5 percent level at several horizons around 13 years after the shock. Overall, the estimates of the TFP response to a defense shock are considerably more uncertain, and they are insignificant at conventional levels for most horizons. Whereas the simple specification shows essentially no impact on TFP at shorter horizons, the baseline specification shows evidence of a positive near-term TFP response to a defense shock, with point estimates that are marginally significant between two and eight years. However, the main conclusion is that—unlike for nondefense R&D—there is 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.

As Figure 6 shows, including the additional controls qualitatively has no major effects on the estimated TFP responses at longer horizons. Appendix C.1 and C.2 show that the TFP responses to both shocks also remain very similar if we use all appropriation changes rather than just those classified as exogenous, or if we use the raw narrative measures rather than the orthogonalized ones. Appendices C.3 and C.4 further document that the significant positive long-run TFP response to a nondefense R&D shock is robust to many different additions to the control set x_t, as well as to various other changes in specification. Together, these results suggest that policy endogeneity is not as serious a concern for government R&D as it is for broader changes in tax or spending policies. Nevertheless, we include four lags of the baseline controls in all remaining specifications, and—unless mentioned otherwise—we continue to use the same orthogonalized narrative measures for identification.

C. Effects on Other Productivity and Innovation Indicators

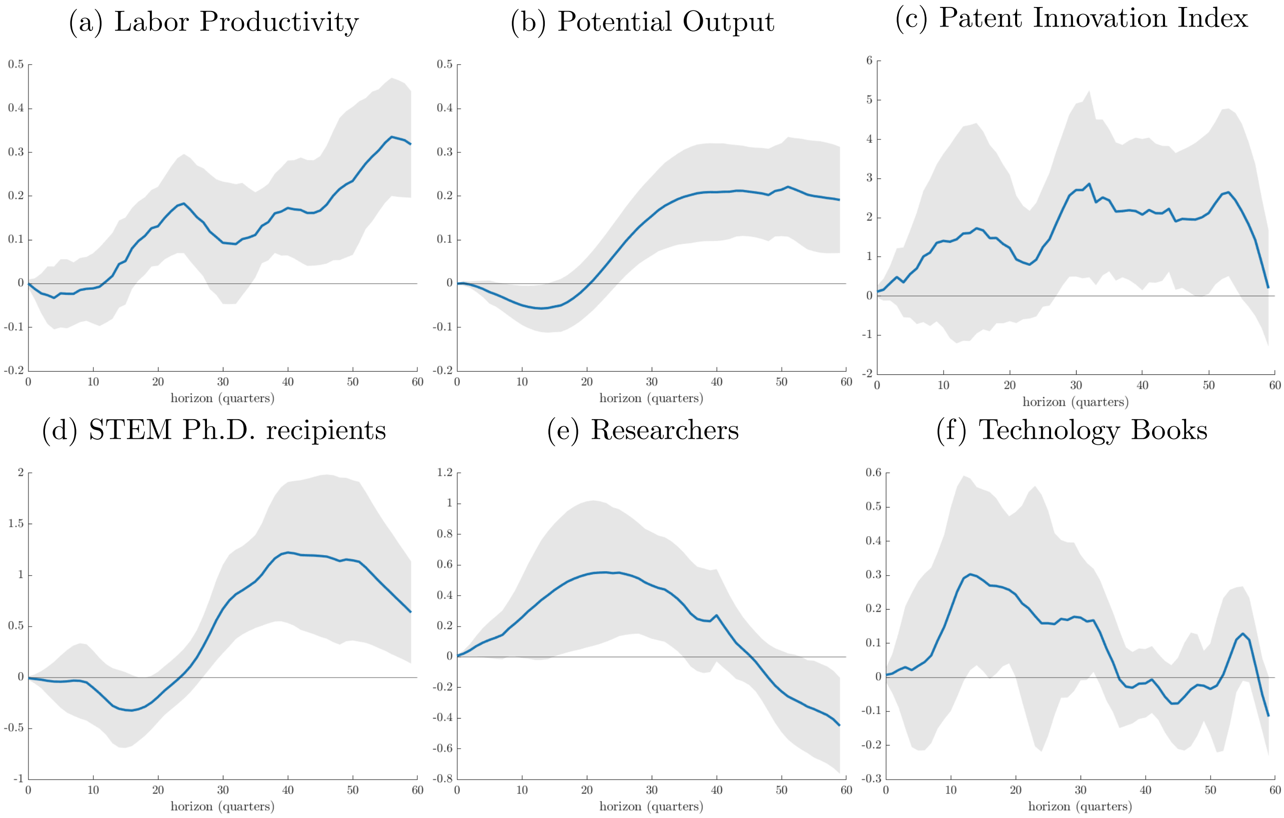
Figure 7 reports impulse responses of several other productivity and innovation indicators to nondefense R&D shocks. The estimates are again based on (2) with the six baseline controls in x_t, and scaled to imply a one percent peak increase in total government R&D capital. For brevity, the corresponding results for defense R&D shocks are reported in Appendix C.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 0.30 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 C.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 around 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 appropriations is not a source of fluctuations at business cycle frequencies. Indeed, Appendix C.8 documents that real GDP shows no short-run response to a nondefense appropriations shock, but simply rises in the longer run along a trajectory that is very similar to that of business-sector labor productivity.

The next panel in Figure 7 shows the response of the Congressional Budget Office (CBO) measure of potential GDP (in logs), 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, however, there is a gradual and significant increase in potential output, which expands 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 an 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), Bluwstein et al. (2020), 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

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



Notes: Estimates based on (2) using the orthogonalized narrative measure of changes in nondefense R&D appropriations, see (1). Lazarus et al. (2018) HAR bands are for 95 percent confidence levels. 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.

the quarterly version constructed by Cascaldi-Garcia and Vukotić (2022), which is 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 temporarily rises by around 2 percent after a positive shock to nondefense R&D appropriations, an increase that is significant at several medium-run horizons. The rise in patent grants with economic value occurs well in advance of the increase in TFP seen in Figure 6 and fades near the end of the 15-year forecast horizon. The timing of the response is consistent with increased government funding for nondefense R&D leading to more patents with economic value followed by related 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 12 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 around 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. After taking into account the additional lags between R&D appropriations and outlays, the timing of the effects also appears broadly in line with existing evidence that innovation and productivity responses typically lag private R&D spending by two- to five years, see e.g., Hall et al. (2010) for an overview.

Appendix C.7 shows that, in contrast to a nondefense shock, a positive shock to defense R&D does not lead to similarly unambiguous increases in the same productivity and innovation indicators, reinforcing our earlier conclusion that defense R&D spending on average does not appear to have the same positive long-run spillovers on business-sector productivity as nondefense R&D within the horizons that we consider.

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

Figure 6 showed that the shocks to appropriations for nondefense and defense R&D lead to hump-shaped increases in government R&D capital. 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 additional information available in 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},

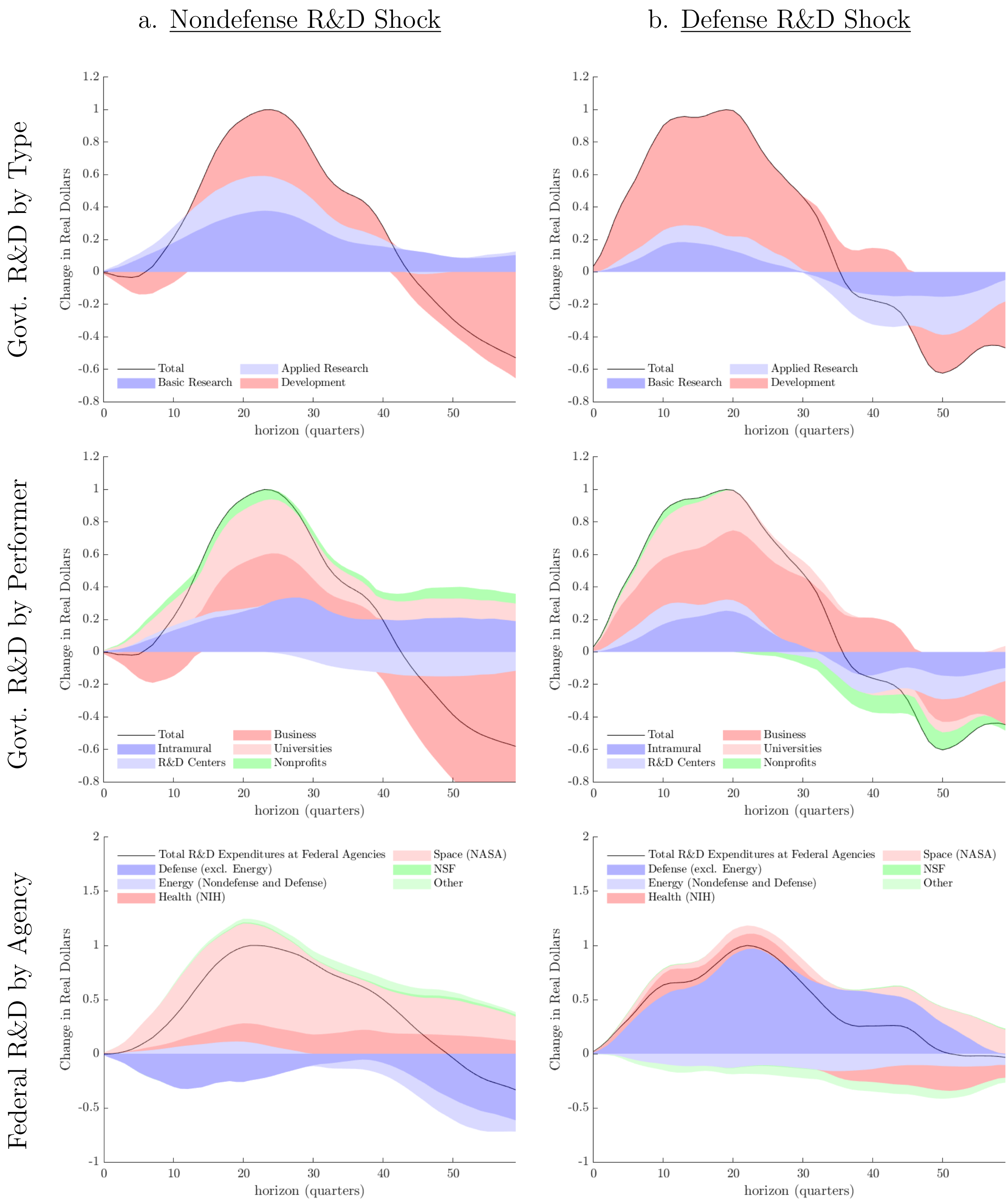
$$\Delta \ln I_t^{tot} \approx \sum_j \frac{s_t^j + s_{t-1}^j}{2} \Delta \ln I_t^j \tag{3}$$

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 (3) using the baseline specification in (2). 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 panel) or defense (right panel) 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 of R&D and by performing sector. As can be seen in the figure, both shocks lead to a gradual build-up in total R&D spending flows, and partial spending reversals in the longer run. In response to a nondefense shock, R&D spending is approximately unchanged for the first six quarters, after which it slowly rises to a peak after about six years, and subsequently gradually declines. After about 10 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 starts immediately on impact.

The decomposition in the first row of Figure 8 shows that both shocks lead to increases in each of the three types of R&D investment (basic research, applied research, and development) during the boom phase. However, the nondefense shock leads to a substantially larger increase in basic and applied research (up to 38 cents and 22 cents, respectively). A defense shock instead leads mostly to increases in spending on development (up to 75 cents). For the nondefense shock, the eventual reversal in spending is exclusively in development, while funding for basic research remains elevated throughout. For the defense shock, the spending reversal is instead 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.

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



Notes: Estimates based on (2) using the orthogonalized narrative measure of changes in federal nondefense (left panel) and defense (right panel) R&D appropriations, see (1). 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.

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 33 cents, while a defense shock raises intramural spending by at most 27 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 exclusively by decreases in funding for businesses and R&D centers. The increase in funding for research at universities and government agencies is instead highly persistent, which likely mirrors the persistent impact of the nondefense shock on funding for more fundamental research. For the defense shock, in contrast, the reversal in spending affects R&D activities by 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, and the NSF. 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 spending also appears to crowd out some funding for defense R&D, as there are decreases in DOD outlays for R&D throughout the entire forecast horizon. Energy R&D spending—which covers both defense and nondefense functions—initially increases, but decreases at longer horizons. Unsurprisingly, the bottom right panel of Figure 8 shows that a positive defense R&D shock mainly leads to DOD spending increases. While the defense shock leads to some persistent declines in R&D outlays for energy and NASA, and at later horizons also for health, there is little evidence of very large crowding-out effects on the nondefense agencies' R&D funding.

The decomposition by federal agency shows that, in dollar terms, the nondefense shock primarily induces a change in R&D funding for NASA. This finding suggests that changes in appropriations for NASA, especially at the time of the agency's rapid growth during the space race, are potentially very important as a source of identifying variation. In Appendix C.4, we show that the positive TFP response to a nondefense shock is robust to excluding NASA appropriations during the height of the space race (1958-63). Whereas the uncertainty around the estimates increases meaningfully, the long-run TFP increase remains statistically significant and similar in size. In Sections 4 and 5 below, we will consistently report results for specifications that omit the height of the space race from the narrative measures to verify the sensitivity of the results to this potentially influential part of the sample.

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},

$$\Delta \ln K_t^{tot} \approx \sum_j \frac{s_t^j + s_{t-1}^j}{2} \Delta \ln K_t^j \tag{4}$$

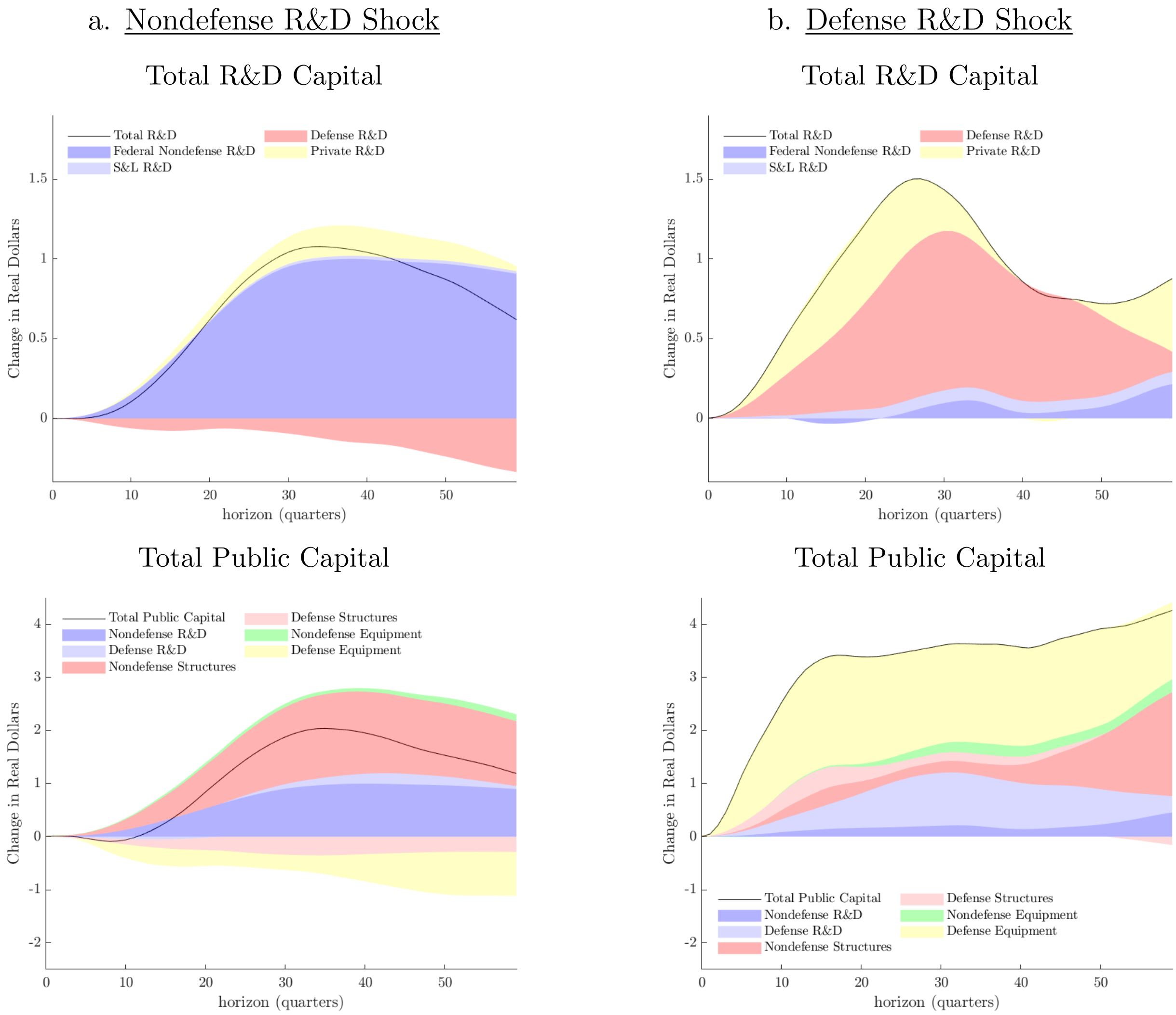
where K^j is R&D capital of category j in constant dollars and s^j denotes the nominal share of capital of category j in total R&D capital ($s^j = K^{n,j}/K^{n,tot}$, where $K^{n,j}$ is capital in category j in current dollars). The four main funders of total R&D are (i) federal defense agencies, (ii) federal nondefense agencies, (iii) state and local governments, and (iv) the private sector. The contributions of each funding category are the cumulative impulse response of the individual terms in (4) estimated with the baseline specification in (2). 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 panels in Figure 9 show that the defense and nondefense appropriations shocks primarily affect R&D capital within their own category. A positive nondefense shock leads to some crowding out of defense R&D capital, by up to 20 cents after 15 years, while there is very little effect of a defense shock on nondefense R&D capital. Consistent with the framework in Akcigit et al. (2020) and the evidence in De Lipsis et al. (2022), there are increases in private R&D capital in response to both the nondefense and defense shocks. The increases in private R&D capital following a nondefense shock, however, are relatively small, peaking at 19 cents per federally funded dollar. For defense R&D, the peak increase in private R&D capital is somewhat larger at 52 cents per federally funded dollar.[12]

To investigate how the shocks to government R&D affect the various other components

[12] Changes in domestic R&D spending may also affect R&D spending in the rest of the world, which in turn could lead to domestic spillovers. Because of data availability, we do not look into the role of such global spillovers.

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



Notes: Estimates based on (2) using the orthogonalized narrative measure of changes in federal nondefense (left panel) and defense (right panel) R&D appropriations, see (1). 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.

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

The bottom left panel in Figure 9 shows that, in the decomposition of total public capital, there is no evidence of the nondefense shock crowding out defense R&D capital. However, the nondefense shock leads to a broader reallocation from defense to nondefense public capital. While the defense capital stocks for structures (e.g., military bases and facilities) and equipment (e.g., ships and aircraft) decline, 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 1.58 dollars after about 8 years. While this increase is much smaller than the average ratio between the nondefense structures and R&D capital stocks, it is almost certainly too large to be explained exclusively by the reclassification of R&D plant expenditures as 'structures' in the national accounts. In Appendix C.9, we present 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. 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, the federal appropriations bills financing increases in R&D generally do not provide significant funding for public infrastructure investment via transfers to state and local governments. In Appendix C.9, we further show that federal transfers, if anything, decline in response to a positive nondefense R&D shock. The rise in investment expenditures by state and local government is instead financed initially by debt, and later on by increases in tax revenues relative to other expenditures. The growth in state and local structures is also broad-based, with the largest increases in education structures (schools and universities), followed by highways and streets, and power, water, and sewer systems. Overall, the expansion in public infrastructure appears large enough to potentially generate meaningful indirect productivity effects.

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 only negligible 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 2.48 dollars), and also a smaller increase in defense structures (up to 48 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.[13] More importantly, the annual DOD appropriations bills that fund defense R&D also fund procurement (i.e., equipment), and funding for developing new military hardware typically leads to subsequent purchases of the newly developed equipment. Moreover, 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. In contrast to nondefense public infrastructure, however, 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

[13]Due to data limitations, our measure of DOD R&D appropriations is based on the full Research, Development, Test, and Evaluation (RDT&E) accounts. Defense structures are largely funded by the separate Military Construction and Veterans Affairs Appropriations bills, not the Defense Appropriations bills funding RDT&E activities.

no effect, e.g., CBO (2016). The increases in non-R&D defense capital following defense R&D shocks are therefore unlikely to be a major influence on the long-run TFP response.

The main conclusions regarding the indirect channels for long-run TFP spillovers are the following: First, the impact of shocks to government R&D in one category (i.e., defense or nondefense) on the other is relatively small or absent. The orthogonalized narrative measures therefore appear largely successful in picking up idiosyncratic changes in nondefense or defense R&D, such that reallocations of resources across both categories of R&D are unlikely to be important for the TFP responses in Figure 6. 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 private R&D capital are relatively small, especially after a nondefense R&D shock. Finally, the responses of public infrastructure capital are sizeable. Given the widespread evidence for productivity spillovers of public infrastructure, these responses are potentially important in determining the long-run TFP responses to the appropriations shocks.

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,

$$\Delta f_t = \alpha_t' \Delta m_t + \eta_t \Delta q_t + \phi_t \Delta k_t + \Delta \nu_t \tag{5}$$

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 ν_t is technological progress after accounting for growth in both types of public capital, q_t and k_t. The parameters in α_t are the production function elasticities for all private inputs, η_t is the elasticity to public infrastructure, and ϕ_t is the elasticity to government R&D capital. As is the convention in the literature, we assume that (non-R&D) defense capital does not generate any TFP spillovers. The notation henceforth assumes that all growth rates are demeaned such that constants are omitted without loss of generality.

Defining $\Delta tfp_t = \Delta f_t - \alpha_t' \Delta m_t + \epsilon_t$, and assuming constant elasticities with respect to public capital, (5) can be rewritten as

$$(6) \qquad \Delta tfp_t \;=\; \eta \Delta q_t + \phi \Delta k_t + \Delta w_t \;\;,\;\; \Delta w_t = \Delta \nu_t + \epsilon_t \;\;,\;\; E[\Delta w_t] = 0$$

where Δtfp_t is the utilization-adjusted measure of business-sector TFP growth (or Solow residual) constructed by Fernald (2012), and ϵ_t is measurement error. The unobserved residual term Δw_t consists of the productivity growth term $\Delta \nu_t$, as well as any discrepancy ϵ_t between measured TFP and actual productivity growth. Apart from measurement errors in Δf_t and Δm_t, the discrepancy between measured and actual productivity growth could be due to the mismeasurement of the elasticities in α_t. As explained in Fernald (2012), the identification of α_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, ϵ_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 η as a known constant, and estimate ϕ across a range of values of η consistent with the empirical literature. A recent survey by Ramey (2021) establishes a plausible range of 0.065 to 0.12 for η. We use these endpoints to estimate a corresponding range for ϕ, 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 η 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 (6), this definition leads to the structural estimation equation,

$$(7) \qquad \Delta \widetilde{tfp}_t \;=\; \phi \Delta k_t + \Delta w_t$$

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 ϕ in (7) 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 D.1. In our application, we use the responses to R&D appropriations shocks discussed in the previous section.[14] Note that the functional form in (7) makes very specific assumptions about the lags between R&D spending and the productivity effects. Appendix D.1 shows

[14] 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 C.4 shows that the impulse response estimates are very similar in the balanced sample to those shown in Figure 6.

that these assumptions in fact align very well with the impulse response estimates, which are obtained without imposing any rigid assumptions about the timing of the effects.

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 (2). 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 ϕ are

$$E\left[w_t^{\perp}(h) z_t^{\perp}\right] = 0 \; ; \; h = 0, \dots, H-1 \quad , \; w_t^{\perp}(h) \equiv \widetilde{tfp}_t^{\perp}(h) - \phi k_t^{\perp}(h) \tag{8}$$

where $z_t^{\perp}$ is the one-step ahead forecast error from the linear projection of z_t on Ω_{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 Ω_{t-1}. Intuitively, the identifying conditions in (8) exploit the fact that, if the structural relationship in (7) holds in the raw data, it also holds across the $h+1$-step ahead forecast errors after projection on Ω_{t-1} for any forecast horizon h. The key exogeneity assumption in (8) is that, after projection on Ω_{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 unpredicted by the projection on Ω_{t-1}, or because of accumulated measurement error ϵ in measured TFP between period t and $t+h$ that is unpredicted by projection on Ω_{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 Ω_{t-1} suffices to remove any joint influences of *past* shocks (realized prior to quarter t) on z_t and future productivity growth.

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 (8) 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 Ω_{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 (8) would not necessarily hold, and the SP-IV estimate of ϕ would be potentially biased. However, the estimated impulse responses of business-sector labor and non-R&D capital inputs to R&D appropriations shocks, reported in Appendix C.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.[15] 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 capital. As shown earlier, both R&D shocks lead to increases in private R&D capital. If the methodology in Fernald (2012) underestimates the aggregate elasticity of private R&D capital, the estimates of ϕ are likely to be biased upward. Fortunately, Figure 9 showed that the increases in business-sector R&D capital are relatively small, especially for the nondefense R&D shock, such that the bias is likely relatively 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 (7) 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. We therefore also consider specifications that allow 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 equation is adjusted as follows:

$$(9) \qquad \Delta \widetilde{tfp}_t \;=\; \phi_{ND}\left(s_{ND,t}\Delta k_t^{ND}\right) + \phi_D(1 - s_{ND,t})\Delta k_t^D + \Delta w_t \quad , \quad E[\Delta w_t] = 0$$

This specification assumes production function elasticities to Δk_t^{ND} and Δk_t^D that scale with $s_{ND,t}$ and $1-s_{ND,t}$, such that ϕ_{ND} (ϕ_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 ϕ_{ND} and ϕ_D can be compared to the estimates of ϕ in the simpler specification in (7). For the purpose of calibrating an aggregate production function with constant elasticities on Δk_t^{ND} and Δk_t^D, the estimates of ϕ_{ND} and ϕ_D can be multiplied by 0.5, which is approximately the average of $s_{ND,t}$ across the sample. An alternative approach, pursued in Appendix D.4,

[15] 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 C.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 ($0.2 \times 0.33 \times 0.15 \approx 0.01$).

treats the elasticities to Δk_t^{ND} and Δk_t^D as constants in the estimation.

When $\phi_{ND} \neq \phi_D$, the estimates of ϕ in the simpler specification in (7) are not necessarily consistent for either ϕ_{ND} or ϕ_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 do not show much crowding-out of one type of government R&D capital by the other, such that we expect both specifications to provide similar estimates.

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 (Han and Phillips 2006; 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 $h = 3, 7, 11, ..., 59$.[16]

B. Estimation Results

Table 1 reports estimates of ϕ, ϕ_{ND}, and ϕ_D for various specifications, together with 95 percent weak-instrument-robust confidence intervals. The first five rows show estimates of ϕ in (7), including only total government R&D capital, whereas the remaining rows show estimates for ϕ_{ND} and ϕ_D in (9), 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 η are omitted.

The first row in Table 1 shows estimates based on the impulse responses identified with the (orthogonalized) narrative measure for nondefense appropriations, z_t^{ND}. For $\eta = 0.08$, the point estimate of ϕ in (7) based on the impulse responses to a nondefense shock is 0.12. This estimate is highly statistically significant and fairly precisely estimated, with a 95 percent robust confidence interval ranging from 0.09 to 0.16. As expected, the point

[16]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	0.12*** (0.09,0.16)		0.12*** (0.10,0.16)	0.11*** (0.08,0.15)
[2]	Total	Exo ND, No Space	0.14*** (0.09,0.32)		0.14*** (0.09,0.33)	0.13*** (0.08,0.30)
[3]	Total	All ND	0.11*** (0.09,0.15)		0.12*** (0.09,0.16)	0.10*** (0.08,0.14)
[4]	Total	Exo D		−0.24 (−1.46,0.02)		
[5]	Total	All D		−0.23 (−1.30,0.03)		
[6]	ND/D	Exo ND	0.11*** (0.06,0.23)	−0.01 (−0.25,0.45)	0.11*** (0.06,0.23)	0.10*** (0.05,0.22)
[7]	ND/D	Exo ND/D	0.10*** (0.06,0.17)	−0.07 (−0.25,0.39)	0.10*** (0.07,0.18)	0.09*** (0.06,0.17)
[8]	ND/D	Exo ND, No Space	0.14 (−2.00†,0.51)	0.18 (−2.00†,2.00†)	0.14 (−2.00†,0.51)	0.13 (−2.00†,0.50)
[9]	ND/D	All ND	0.11*** (0.06,0.21)	−0.02 (−0.25,0.40)	0.11*** (0.06,0.21)	0.10*** (0.05,0.19)

Notes: Rows [1]-[5], SP-IV estimates of ϕ (government R&D) in (7); rows [6]-[9], SP-IV estimates of ϕ_{ND} (nondefense R&D) and ϕ_D (defense R&D) in (9). All specifications include the baseline set of lagged controls described in Section 3. Numbers in parentheses are 95 percent weak-instrument-robust confidence intervals based on inverting the KLM statistic of Kleibergen (2005). Test inversion is limited to a grid with endpoints −2 and 2, † denotes intervals constrained at these endpoints. Subvector inference in rows [6]-[9] is based on the projection method. Stars $*$, $**$ and $***$ denote statistical significance at 10, 5 and 1 percent levels, respectively. 'Exo ND/D' denotes the orthogonalized narrative measure of exogenous changes in nondefense/defense R&D appropriations. 'All ND/D' denotes the orthogonalized series of all changes in nondefense/defense R&D appropriations, ignoring our narrative classification. 'No Space' indicates that the instrument is also orthogonalized to all changes in space appropriations between 1958 and 1963. Sample: 1948Q1–2021Q4.

estimates are decreasing in the assumed value of η, with $\hat{\phi} = 0.12$ for $\eta = 0.065$ and $\hat{\phi} = 0.11$ for $\eta = 0.12$. 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 D.1. Consequently, the increase in TFP after adjusting for public infrastructure is smaller when η is larger. However, in practice the estimates of ϕ are very similar across Ramey's (2021) plausible range of values for $\eta \in [0.065, 0.12]$.[17]

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 is further orthogonalized to all changes in appropriations for NASA from 1958 to 1963, the period with the fastest growth in nondefense government R&D capital in the sample. The resulting point estimates remain highly significant and

[17] The point estimate is $\hat{\phi} = 0.16$ when assuming $\eta = 0$, and $\hat{\phi} = 0.04$ when $\eta = 0.39$. The latter value is based on Aschauer (1989) and is the highest estimate mentioned in Ramey (2021).

are slightly larger than in row [1], around 0.13 to 0.14 depending on η. Without the space race as identifying variation, the robust confidence intervals become notably wider, and in particular, substantially larger values of ϕ cannot be ruled out. Row [3] shows estimates based on responses to all changes in nondefense R&D appropriations, after orthogonalizing to all defense R&D changes, regardless of their narrative classification. The estimates are very similar to those in row [1], and the narrative classification therefore matters little for the identification of ϕ.

The next two rows in Table 1 report estimates of ϕ identified with impulse responses to defense R&D shocks rather than nondefense shocks. Row [4] reports $\hat{\phi} = -0.24$ based on the (orthogonalized) narrative measure of exogenous changes in defense R&D appropriations z_t^D, and row [5] shows that $\hat{\phi} = -0.23$ when using all changes in defense R&D appropriations regardless of their narrative classification. Unlike for the nondefense R&D shocks, both estimates are negative. The confidence bands are very wide, and none of the estimates are significant at conventional levels. The narrative classification is again unimportant.

The remaining rows in Table 1 report estimates of ϕ_{ND} and ϕ_D from the specification in (9) that includes both types of government R&D capital simultaneously, with subvector inference based on the projection method. In row [6], ϕ_{ND} and ϕ_D are identified jointly using the same narrative nondefense measure as in row [1]. The resulting point estimate of ϕ_{ND} is 0.11 for the intermediate and low values of η, and 0.10 for the high value of η, all of which are statistically significant and close to the corresponding estimates in row [1]. In contrast, the point estimate of ϕ_D is small, -0.01, and statistically insignificant.

Rows [7]-[9] in Table 1 provide additional estimates of ϕ_{ND} and ϕ_D identified with different impulse responses. In row [7], identification is based on impulse responses to both defense and nondefense R&D shocks using the two original exogenous narrative measures $\Delta a_t^{exo,i}/K_{t-4}^i, i = D, ND$, i.e., without mutual orthogonalization. Rows [8] and [9] are instead based on the same narrative measures of nondefense R&D shocks as in rows [2] and [3], i.e., excluding the space race and using all changes in nondefense R&D appropriations, respectively. The estimates of ϕ_D range from -0.07 to 0.18, but are statistically insignificant in all specifications. The estimates of ϕ_{ND}, on the other hand, are close to those in rows [2] and [3], and remain highly statistically significant. The only exception is in row [8]: Without the large NASA appropriations during the space race, 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 rejection of no spillovers even when excluding the space race, see Appendix D.3. For the interested reader, the same Appendix provides further robust inference results, including the simultaneous confidence sets associated with the estimates in rows [6]-[9].

Appendix D.4 reports results for the alternative version of (9) that instead assumes constant elasticities to Δk_t^{ND} and Δk_t^{D}. The results are broadly consistent with those in Table 1. After scaling appropriately for comparability, the estimates of ϕ_{ND} range from 0.06 to 0.13, but are generally somewhat smaller than those reported in Table 1. The estimates of ϕ_D for the alternative specification range from -0.09 to 0.16, and are all insignificant and imprecisely estimated.

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 0.14, with a midpoint of approximately 0.12. Multiplying by the average postwar share of nondefense R&D capital of 0.5, the estimates imply elasticities to nondefense R&D capital ranging from 0.045 to 0.07, with a midpoint of 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 sizeable direct spillovers of nondefense government R&D on business-sector productivity. In contrast, the elasticity estimates based on variation in defense R&D vary considerably across specifications, from -0.24 to 0.18, and come with wide confidence bands. Unlike for nondefense R&D capital, we cannot draw any sharp conclusions regarding the size—or even the sign—of any direct spillovers of defense R&D.

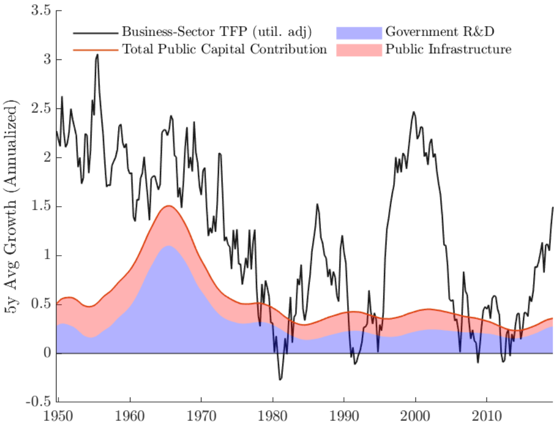
V. The Macroeconomic Returns to Government R&D

A. Historical Contributions to TFP Growth

With the estimates of the TFP spillovers of government R&D in hand, it is possible to assess the contribution of public capital accumulation to postwar business-sector TFP growth. 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, for each of the three different values of η. The contribution of public infrastructure is calculated as $\eta\Delta q_t$. The figure in the left panel of Figure 10 shows the resulting contributions of government R&D and public infrastructure for $\eta = 0.08$. The table in the right panel of Figure 10 reports averages over selected time windows for each of the three values of η.

The main finding is that government R&D has contributed substantially to total TFP growth since WWII R&D—accounting for roughly one-quarter of the total, on average—regardless of the value of η within Ramey's (2021) plausible range. The contribution of government R&D is consistently similar in size to the contribution of public infrastructure

Figure 10: 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 η			
Infrastructure	0.33	0.19	0.19	0.09
R&D	0.53	0.27	0.22	0.21
	b. Low η			
Infrastructure	0.27	0.16	0.15	0.07
R&D	0.55	0.28	0.22	0.22
	c. High η			
Infrastructure	0.50	0.29	0.28	0.14
R&D	0.49	0.25	0.20	0.20

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 η.

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 0.82 to 0.99 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 about twice as large as that of public infrastructure: 0.55 versus 0.27 percentage points, respectively. For the high value $\eta = 0.12$, each component of public capital contributes about half a percentage point. 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 0.38 to 0.45 percentage points as η increases from low to high. For low η, around 70 percent of the contribution of public capital $((0.55 - 0.28)/0.38 = 0.71)$ is due to the slowdown in government R&D. For high η, the slowdown in R&D contributes slightly more than half $((0.49 - 0.25)/0.45 = 0.53)$. According to our estimates, therefore, the retrenchment of government R&D in the 1970s and 1980s was at least as important for explaining the slowdown in productivity as the slower pace of public infrastructure investment. The contribution of government R&D to TFP growth is 20 to 22 basis points in both 1990-2009 and 2010-2021. In contrast, the contribution of public infrastructure to TFP growth fell by half in 2010-2021 relative to 1990-2009 as a result of the further slowing of growth in public infrastructure.

The left panel in Figure 10 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 η 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 figure in the left panel also shows that public investment—either in R&D

or infrastructure—plays little role in accounting for the high TFP growth immediately after WWII. It is possible that the higher TFP growth in the 1950s was partially driven by wartime defense R&D—see, for instance, Gross and Sampat (2023)— which plays no role in the decomposition because of our assumption that $\phi_D = 0$.[18] Finally, government R&D also matters little for the TFP burst during the IT revolution in the 1990s.

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 δ 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 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 an 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 ϕ, which is assumed to be constant over the estimation sample. A common alternative approach to estimating returns is instead to estimate ρ 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 (7) yields

$$\Delta \widetilde{tfp}_t \quad = \quad \rho \frac{\Delta K_t}{Y_t} + \Delta w_t \tag{10}$$

To estimate ρ, 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 (10) 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:

$$\Delta \widetilde{tfp}_t \quad = \quad \rho_{ND} \frac{\Delta K_t^{ND}}{Y_t} + \rho_D \frac{\Delta K_t^D}{Y_t} + \Delta w_t \tag{11}$$

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.

[18] Including wartime R&D could offer more identifying variation for estimating ϕ^D, but is unlikely to influence the estimates of ϕ^{ND}, as federal R&D expenditures were almost exclusively for defense activities before the 1950s.

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

	Government R&D Measure	Instruments	Intermediate $\eta = 0.08$: $\hat{\phi}_{ND} \times \frac{Y}{K}$	Intermediate $\eta = 0.08$: $\hat{\rho}_{ND}$	Low $\eta = 0.065$: $\hat{\phi}_{ND} \times \frac{Y}{K}$	Low $\eta = 0.065$: $\hat{\rho}_{ND}$	High $\eta = 0.12$: $\hat{\phi}_{ND} \times \frac{Y}{K}$	High $\eta = 0.12$: $\hat{\rho}_{ND}$
[1]	Total	Exo ND	2.04	2.13*** (1.32,2.75)	2.11	2.20*** (1.37,2.79)	1.86	1.96*** (1.16,2.65)
[2]	Total	Exo ND, No Sp.	2.43	3.17*** (1.42,5.00†)	2.50	3.23*** (1.50,5.00†)	2.25	3.01** (1.21,5.00†)
[3]	Total	All ND	1.96	1.95*** (1.28,2.54)	2.03	2.01*** (1.34,2.58)	1.78	1.79*** (1.13,2.44)
[4]	ND/D	Exo ND	1.91	2.49*** (0.76,3.95)	1.98	2.55*** (0.82,3.98)	1.74	2.33*** (0.61,3.87)
[5]	ND/D	Exo ND/D	1.67	2.04** (0.13,3.78)	1.73	2.10** (0.17,3.82)	1.50	1.88** (0.01,3.69)
[6]	ND/D	Exo ND, No Sp.	2.42	3.98 (−2.00†,5.00†)	2.49	4.02 (−2.00†,5.00†)	2.24	3.87 (−2.00†,5.00†)
[7]	ND/D	All ND	1.87	2.04*** (0.70,3.68)	1.94	2.10*** (0.75,3.71)	1.70	1.89*** (0.56,3.60)

Notes: Rows [1]-[3], SP-IV estimates of ρ (government R&D) in (10); rows [4]-[7], SP-IV estimates of ρ_{ND} in (11). All specifications include the baseline set of lagged controls described in Section 3. Numbers in parentheses are 95 percent weak-instrument-robust confidence intervals based on inverting the KLM statistic of Kleibergen (2005). Test inversion is limited to a grid with endpoints −2 and 5, † denotes intervals constrained at these endpoints. Subvector inference in rows [4]-[7] is based on the projection method. Stars $*$, $**$ and $***$ denote statistical significance at 10, 5 and 1 percent levels respectively. 'Exo ND/D' denotes the orthogonalized narrative measure of exogenous changes in nondefense/defense R&D appropriations. 'All ND/D' denotes the orthogonalized series of all changes in nondefense/defense R&D appropriations, ignoring our narrative classification. 'No Space' indicates that the instrument is also orthogonalized to all changes in space appropriations between 1958 and 1963. Sample: 1948Q1–2021Q4.

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. The various rows in the table mirror the specifications in Table 1, with rows [1]-[3] reporting results for (10) and rows [4]-[7] reporting results for (11). Identification is based on the same variants of the instruments as in Table 1, and each row also reports the calculation of the rate of return based on the corresponding elasticity estimate in Table 1. The implied net returns can be obtained by subtracting $\delta \approx 0.16$, which is approximately the average depreciation rate for government R&D calculated by the BEA.[19]

As Table 2 shows, the implied rates of return to nondefense R&D are high. The estimates range from around 150 percent to more than 300 percent depending on the specification, the assumed value of η, and the method of calculation. The SP-IV estimates of ρ are highly statistically significant regardless of the value of η. As with the elasticity estimates, the only exception is the specification with both government R&D types and the narrative measure that excludes the space race as the instrument, see row [6]. For lower values of η, more

[19] Appendix D.5 shows that the estimated net returns are increasing in the assumed depreciation rate. The reason is that higher depreciation rates also lower the estimates of the government R&D capital stock. For the specification in row [1] of Table 2, the estimated net returns vary from 125 percent assuming zero depreciation to 237 percent when doubling the depreciation rates assumed by the BEA.

of the TFP increase is attributed to R&D as opposed to public infrastructure, and the estimated returns are therefore decreasing in η. However, the returns do not vary greatly across the plausible range for η within each specification. The estimated returns are also roughly the same regardless of whether they are derived from the elasticity estimates or estimated directly.

An implication of the large returns on government R&D is that there is 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. The estimates also suggest that government funding of nondefense R&D is self-financing from the perspective of the federal budget, at least in the long run. Assuming a return of 200 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 7.5 cents ($\delta/\rho = 0.15/2 = 0.075$), which is substantially below the historical ratio of federal tax revenues to GDP.

As mentioned in the introduction, the existing literature often estimates returns on private R&D that well exceed the returns on other investments. In their survey of firm and industry regression evidence, Hall et al. (2010) conclude that rates of return on private R&D are likely in the range of 20 to 30 percent, though some estimates are as high as 75 percent. These estimates usually do not aim to capture all possible spillovers across firms and industries. In that sense, our relatively higher estimates at the aggregate level are perhaps not too surprising. In a stylized framework, Jones and Summers (2020) calculate an average social rate of return on total R&D expenditures of 67 percent based on aggregate U.S. data. Different from Jones and Summers (2020), but like most others in the literature, our estimates of the rate of return rely on functional form (Cobb-Douglas) assumptions about the aggregate production function that may not be realistic. Nevertheless, our evidence based on appropriations shocks suggests that the return on R&D funded by federal agencies may be significantly greater than 67 percent. As discussed earlier, one plausible explanation is that this funding is more directed towards fundamental research with larger knowledge spillovers, as in the framework of Akcigit et al. (2020). An important implication is that federal R&D policy should not be restricted to tax credits and subsidies for private businesses, but also provide adequate resources for R&D funding by federal agencies.

VI. Avenues for Future Research

This paper contributes new time series 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, 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 (Oct. 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 (2022). *The Long-Run Effects of Government Spending.* Discussion Paper 15062. Centre for Economic Policy Research.

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 (June 2018). "Public R&D Investments and Private-sector Patenting: Evidence from NIH Funding Rules". In: *The Review of Economic Studies* 86.1, pp. 117–152.

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.

Bluwstein, Kristina, Sinem Hacioglu Hoke, and Silvia Miranda-Agrippino (2020). *When Creativity Strikes: News Shocks and Business Cycle Fluctuations*. Discussion Paper 15062. Centre for Economic Policy Research.

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 (2022). "When Does Government Spending Matter? Evidence from a New Measure of Spending".

Cascaldi-Garcia, Danilo and Marija Vukotić (Jan. 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.

— (Aug. 2021). *Effects of Physical Infrastructure Spending on the Economy and the Budget Under Two Illustrative Scenarios*. Tech. rep. Congressional Budget Office.

Cloyne, James (2013). "Discretionary Tax Changes and the Macroeconomy: New Narrative Evidence from the United Kingdom". In: *American Economic Review* 103.4, pp. 1507–28.

Cloyne, James, Joseba Martinez, Haroon Mumtaz, and Paolo Surico (2022). *Short-Term Tax Cuts, Long-Term Stimulus*. Working Paper 30246. National Bureau of Economic Research.

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.

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". mimeo.

Fieldhouse, Andrew J, Karel Mertens, and Morten O. Ravn (Jan. 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.

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.
Han, Chirok and Peter C. B. Phillips (2006). “GMM with Many Moment Conditions”. In: *Econometrica* 74.1, pp. 147–192.
Ilzetzki, Ethan (2022). *Learning by necessity: Government demand, capacity constraints, and productivity growth.* Discussion Paper 17803. Centre for Economic Policy Research.
Jinnai, Ryo (2014). “R&D Shocks and News Shocks”. In: *Journal of Money, Credit and Banking* 46.7, pp. 1457–1478.
Jones, Benjamin H. and Lawrence H. Summers (2020). *A Calculation of the Social Returns to Innovation.* Working Paper 27863. National Bureau of Economic Research.
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 Whalley (2022). *Moonshot: Public R&D and Growth.* mimeo. University of Calgary.
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.
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.
Montiel Olea, José Luis and Mikkel Plagborg-Møller (2021). “Local Projection Inference Is Simpler and More Robust Than You Think”. In: *Econometrica* 89.4, pp. 1789–1823.
Moretti, Enrico, Claudia Steinwender, and John Van Reenen (Nov. 2019). *The Intellectual Spoils of War? Defense R&D, Productivity and International Spillovers.* Working Paper 26483. National Bureau of Economic Research.
Munnell Alicia, H. (Jan. 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.
NSF (Feb. 2022). *National Patterns of R&D Resources: 2019–20 Data Update.* Tech. rep. National Science Foundation.
Plagborg-Møller, Mikkel and Christian K. Wolf (2021). “Local Projections and VARs Estimate the Same Impulse Responses”. In: *Econometrica* 89.2, pp. 955–980.

Ramey, Valerie A. (Feb. 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 Investmentn*. 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.
— (2023). "Presidential Address: Does Monetary Policy Matter? The Narrative Approach after 35 Years". In: *American Economic Review* 113.6, pp. 1395–1423.

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

Andrew J. Fieldhouse Karel Mertens

ONLINE APPENDIX

Contents

A Data Sources and Definitions

Main data sources:

- F-TFP: FRB San Francisco Total Factor Productivity, see also Fernald (2012)
- BEA-NIPA: U.S. Bureau of Economic Analysis National Income and Product Accounts

- BEA-FA: U.S. Bureau of Economic Analysis Fixed Assets Accounts Tables
- NCSES: National Center for Science and Engineering Statistics,
  - National Patterns of R&D Resources
  - Survey of Federal Funds for Research and Development, pre-1999 data from the NCSES/NSF archives

All additions and subtractions involving quantities in chained dollars are based on the Divisia index approximation to chained aggregates, see Whelan (2002). All real quantities are expressed in 2012 dollars using implicit deflators.

Capital stock variables: Quarterly real capital stocks are valued at real cost and constructed using the perpetual inventory method using quarterly NIPA data on real investment and initial capital stocks (year-end 1946) from the BEA-FA tables. Depreciation rates are quarterly interpolations of annual depreciation rates in the BEA-FA tables.

- **Government R&D Capital**: Chained sum of (i) federal nondefense R&D capital stock, (ii) federal defense R&D capital stock, and (iii) state & local R&D capital stock. R&D capital includes the BEA-NIPA categories 'research and development' and 'software development'. Investment series are lines 22, 30, and 38 in BEA-NIPA Table 3.9.3 (converted to 2012 dollars using Table 3.9.5). Depreciation rates are lines 35, 52, and 72 in BEA-FA Table 7.4 (converted to 2012 dollars using Table 7.3) divided by prior year capital stocks in the same lines of BEA-FA Table 7.2 (converted to 2012 dollars using Table 7.1). **Government Nondefense R&D Capital** and **Government Defense R&D Capital** are constructed analogously using the relevant subcategories.
- **Public Infrastructure Capital**: Chained sum of structures and equipment capital stocks for (i) federal nondefense and (ii) state & local governments. Investment series are lines 28, 29, 36, and 37 in BEA-NIPA Table 3.9.3 (converted to 2012 dollars using Table 3.9.5). Depreciation rates are lines 39, 40, 56, and 57 in BEA-FA Table 7.4 (converted to 2012 dollars using Table 7.3) divided by prior year capital stocks in the same lines of BEA-FA Table 7.2 (converted to 2012 dollars using Table 7.1).
- **Defense Capital**: Chained sum of defense structures and defense equipment capital stocks. Investment series are lines 20 and 21 in BEA-NIPA Table 3.9.3 (converted to 2012 dollars using Table 3.9.5). Depreciation rates are lines 23 and 30 in BEA-FA Table 7.4 (converted to 2012 dollars using Table 7.3) divided by prior year capital stocks in the same lines of BEA-FA Table 7.2 (converted to 2012 dollars using Table 7.1).
- **Business-Sector R&D Capital**: Aggregate of BEA-NIPA categories 'research and development' and 'software development' for the business sector based on the weights and growth rates in F-TFP (`'wgt_r_and_d'`,`'dk_r_and_d'`,`'wgt_software'`, and `'dk_software'`), cumulated and converted to 2012 dollars using BEA-FA Table 7.1.

- **Total R&D Capital**: Chained sum of the components of government R&D capital and business-sector R&D capital.
- **Total Public Capital**: Chained sum of the components of government R&D capital, public infrastructure capital and defense capital.

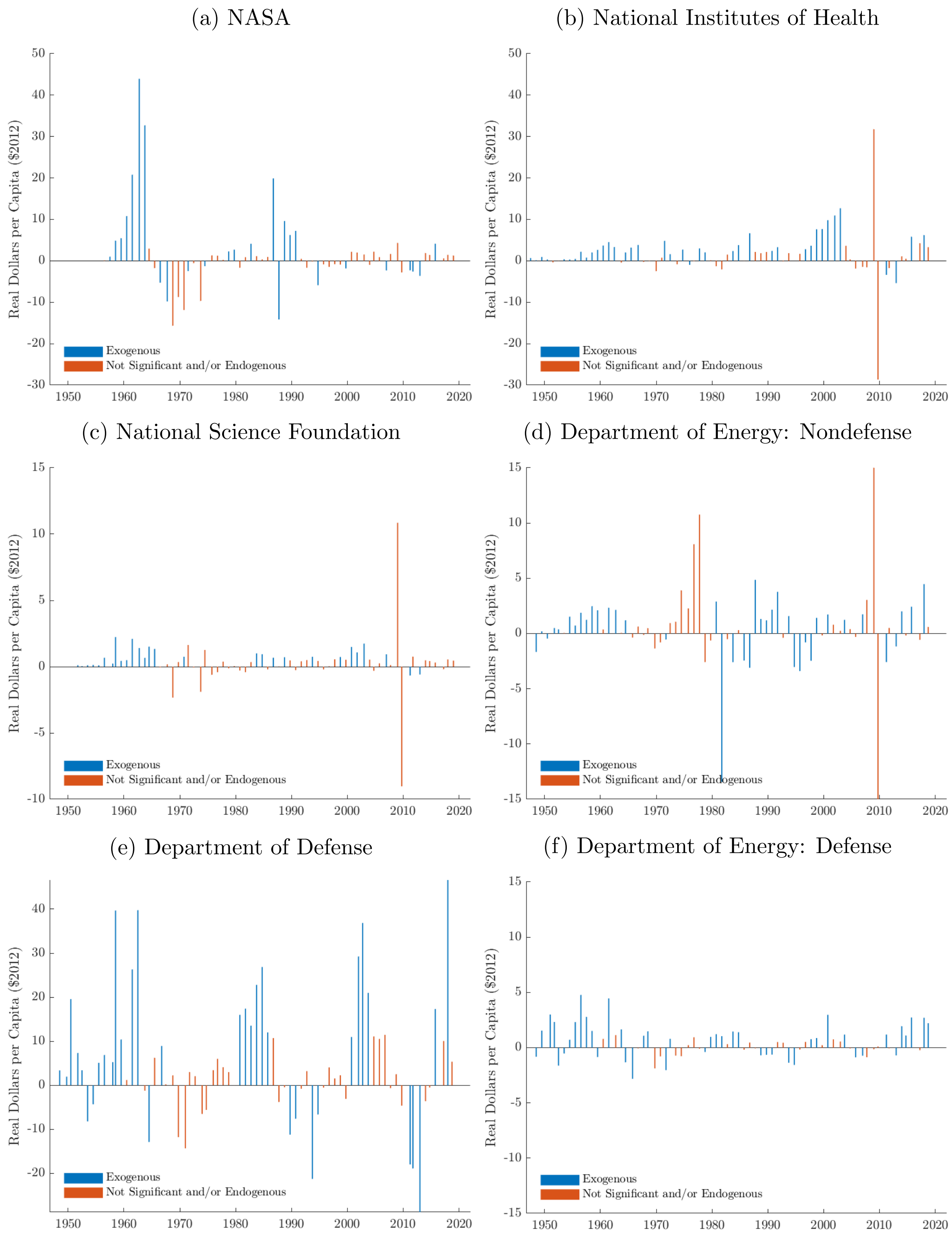
Other variables:

- Variables from F-TFP: **Business-Sector TFP**: utilization-adjusted total factor productivity (F-TFP: `'dtfp_util'`); **Capacity utilization**: (F-TFP: `'dutil'`); **Labor Productivity**: (F-TFP: `'dLP'`); Log-level variables are obtained as cumulative sums of the annualized growth rates in the F-TFP dataset after dividing by 400.
- **Potential Output**: CBO estimate of potential real GDP. From 1949Q1 onward, `'GDPPOT'` from FRED. Observations before 1949Q1 are from the replication files of Ramey and Zubairy (2018).
- **Stock market returns**: Average of the cumulative sums of the equally weighted returns for manufacturing (`'R_EW_Manuf'`), high tech (`'R_EW_HiTec'`), and health industries (`'R_EW_Hlkth'`) from the Kenneth French Data Library (5 Industry Portfolios).
- **Military News**: `'news'` in replication files of Ramey and Zubairy (2018) converted to 2012 dollars by the implicit GDP deflator, divided by potential output.
- **Patent Innovation Index**: Quarterly version of the patent innovation index of Kogan et al. (2017), from the replication files of Cascaldi-Garcia and Vukotić (2022).
- **New PhDs in STEM**: Total number of doctoral recipients in science and engineering. Data for 1947-1957 is from the Historical Statistics of the U.S. (Colonial Times to 1970), series H766-787. Data from 1958 onward is from the NCSES Survey of Earned Doctorates. Quarterly interpolation of annual data.
- **Researchers**: Total researchers (full-time equivalents), from the OECD Main Science and Technology Indicators. Pre-2000 data is obtained from the replication files of Bloom et al. (2020). Quarterly interpolation of annual data.
- **Technology Books**: Books published in the field of technology, constructed Alexopoulos (2011) and obtained from the replication files of Kogan et al. (2017). Quarterly interpolation of annual data.

B Narrative Appropriations Shocks by Agency

Figure B.1 depicts the narrative R&D appropriations changes separately for each agency, before aggregation to nondefense versus defense R&D policy changes, as depicted in Figure 5 of the main text. The top four panels of Figure B.1 depict the R&D appropriations shocks for nondefense expenditures, with NASA in panel (a), NIH in panel (b), NSF in panel (c),

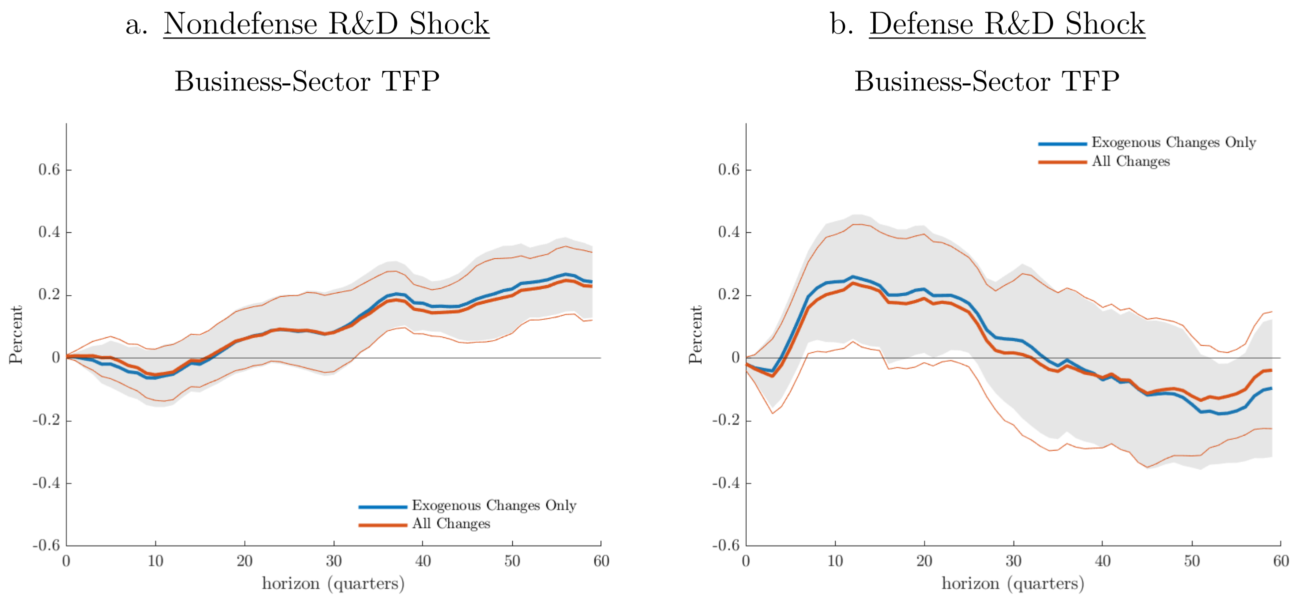
Figure B.1: Changes in R&D Appropriations by Federal Agency



Notes: See Fieldhouse and Mertens (2023). Sample: 1947Q1–2019Q4.

and the nondefense functions of DOE in panel (d). The bottom two panels depict the

Figure C.1: Role of Narrative Classification



Notes: Estimates based on (2) using the measures of changes in federal nondefense (left panel) and defense (right panel) R&D appropriations, see (1). 'Exogenous Changes Only' uses the orthogonalized narratively identified measures as in the baseline specification described in the main text. 'All Changes' uses orthogonalized measures based on all changes in appropriations. Lazarus et al. (2018) HAR bands are for 95 percent confidence levels. Impulses are scaled to imply a 1 percent peak increase in government R&D capital. Sample: 1948Q1–2021Q4.

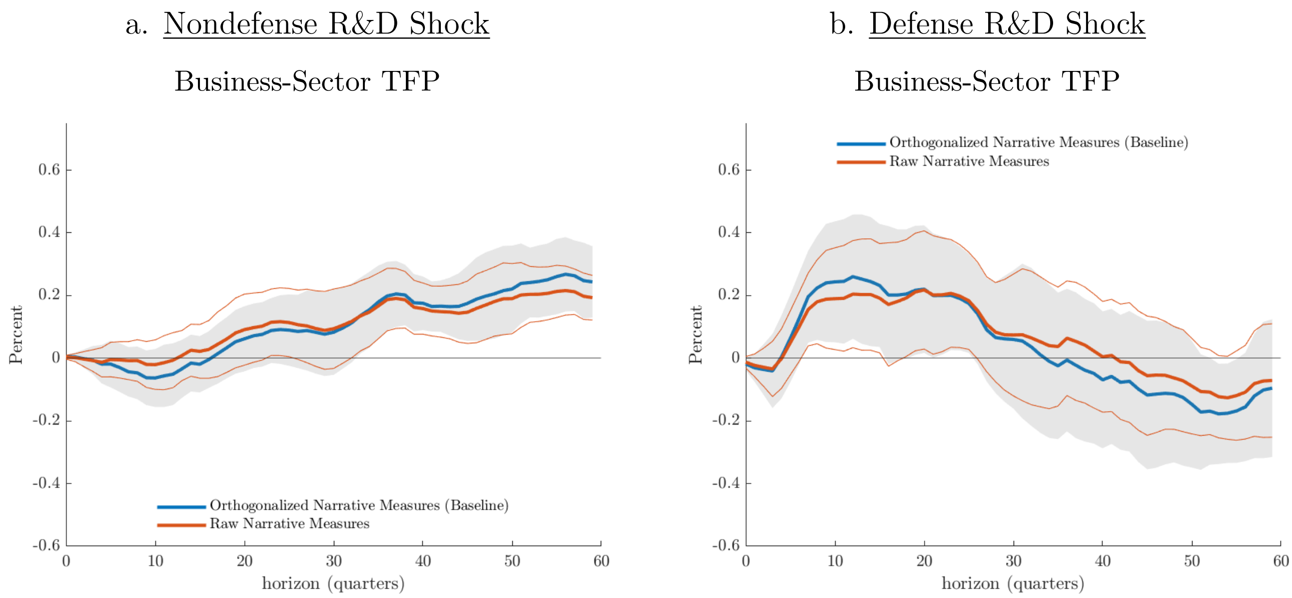
R&D appropriations shocks for defense expenditures, with DOD in panel (e) and the nuclear security functions of DOE in panel (f). Appropriations shocks classified as exogenous are depicted in blue, and those classified as endogenous (or too small to classify) are in red; all R&D appropriations shown are measured in real dollars per capita.

C Impulse Responses: Robustness and Additional Results

C.1 Robustness: Role of the Narrative Identification Step

This section discusses the role of the narrative classification of the changes in federal R&D appropriations as 'exogenous' or 'endogenous' for the impulse response estimates. Figure C.1 replicates the baseline impulse responses of TFP to nondefense and defense shocks from Figure 6 in the main text. The figure also shows estimates for the same specifications, but using all changes in R&D appropriations rather than just those identified as 'exogenous' in the narrative analysis. In this case, the z_t^i variables in (2) contain all changes in appropriations shown in Figure 5, after orthogonalizing defense to nondefense appropriations and vice versa, as in (1). Both the point estimates and confidence intervals for the TFP responses to both the defense and nondefense R&D shocks are very similar when additionally using the endogenous and smaller, unclassified changes in appropriations in the regressions.

Figure C.2: Role of Orthogonalization of the Narrative Measures



Notes: Estimates based on (2) using the measures of changes in federal nondefense (left panel) and defense (right panel) R&D appropriations. Lazarus et al. (2018) HAR bands are for 95 percent confidence levels. Impulses are scaled to imply a 1 percent peak increase in government R&D capital. Sample: 1948Q1–2021Q4.

C.2 Robustness: Role of the Orthogonalization

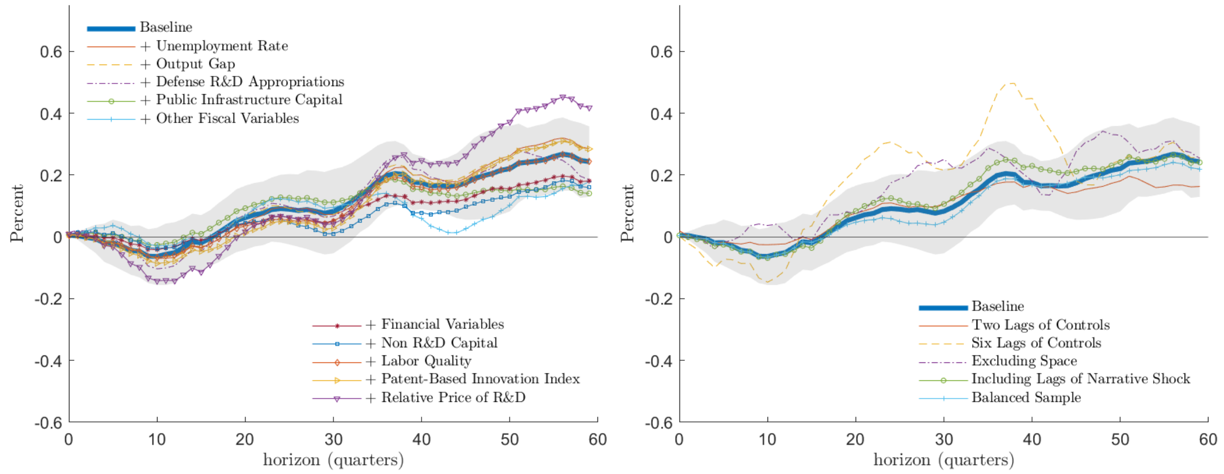
This section discusses the role of mutually orthogonalizing the narrative measures of exogenous changes in defense and nondefense R&D appropriations for the impulse response estimates, as in equation (1) in the main text. Figure C.2 replicates the baseline impulse responses of TFP to nondefense and defense shocks from Figure 6 in the main text. The figure also shows estimates for the same specifications, but using all the raw changes in R&D appropriations $\Delta a_t^{exo,i}/K_{t-4}^i$, $i = D, ND$ as the z_t^i in the local projections in (2) rather the residuals in (1). As the figure shows, the point estimates and confidence intervals for the TFP responses to both the defense and nondefense R&D shocks are very similar.

C.3 Robustness: Additional Control Variables

Figure 6 in the main text shows that including lags of the baseline set of controls x_t reduces the variance of the impulse response estimates to a nondefense R&D shock, but has otherwise no major qualitative effects on the point estimates. This suggests that the controls do not capture any important simultaneous influences on both the narrative measures and future TFP that would threaten the causal interpretation of the estimates in the simpler specification. Here, we consider a number of additions to the baseline set of controls to gain further confidence in the causal interpretation of the positive TFP response to nondefense R&D shocks. Panel (a) of Figure C.3 plots the impulse responses of business-sector TFP

Figure C.3: TFP Impact of Nondefense R&D Shock, Robustness

(a) Additional Control Variables (b) Model Specification



Notes: Estimates based on (2) using the narrative measure of federal nondefense R&D appropriations. Lazarus et al. (2018) 95 percent HAR confidence bands are for the baseline impulse responses. Impulses scaled to imply a 1 percent peak increase in government R&D capital. Sample: 1948Q1–2021Q4 (specification with patent-based innovation index, 1949Q1–2010Q4).

to nondefense R&D shocks for these various additions. For reference, the figure repeats the baseline estimates and the associated 95 percent confidence bands from Figure 6 in the main text. Rows [2]-[11] in Table C.1 report the impulse response coefficients at horizons of 5, 10, and 15 years with HAR confidence bands in parentheses.

As mentioned in the main text, the baseline controls include capacity utilization to capture possible business cycle influences. The first two expanded control sets each add an alternative cyclical indicator: The headline unemployment rate or the output gap (the percentage difference between real GDP and CBO potential output). Neither one has much effect on the estimated TFP response to a nondefense R&D shock, and the TFP response remains highly statistically significant at longer horizons (see rows [2]-[3] in Table C.1). Replacing the utilization rate with either of these alternative cyclical indicators or adding them both at the same time similarly has no major effect on the estimates (these results are not reported).

It is possible that R&D appropriations, despite accounting for only a small share of the federal budget, are predictable by other tax and spending policies that may have independent long-run effects on productivity. For instance, Antolin-Diaz and Surico (2022) find that government spending shocks raise long-run TFP, Cloyne et al. (2022) find that temporary tax cuts have long-run effects on TFP, and Croce et al. (2019) find that the public debt-to-GDP ratio significantly influences the cost of capital for R&D-intensive firms and productivity growth. The baseline controls include lags of cumulative nondefense appropriations, government R&D capital, and the Ramey and Zubairy (2018) military spending news variable. As these variables may not be sufficient to capture all relevant information about fiscal policy, the next three expanded control sets add information about fiscal policy. In turn, we add log cumulative appropriations for defense R&D, the log of the public infrastructure capital stock, and a set of broader fiscal policy indicators. The latter includes the log of total real government consumption expenditures, the ratio of government debt to GDP (based on the Market Value of U.S. Government Debt constructed by the Federal Reserve Bank of Dallas), and the measures of average federal personal and corporate income tax rates from Mertens and Ravn (2013). The addition of defense appropriations has no major impact on the estimates, and the TFP response remains highly statistically significant (row [4] in Table C.1). Adding public infrastructure capital induces a more front-loaded TFP response that is somewhat more muted at longer horizons. The TFP response remains highly statistically significant at longer horizons (see row [5] in Table C.1). Controlling for lags of a broader set of fiscal policy indicators also leads to somewhat smaller TFP responses at longer horizons, but they nevertheless remain highly significant (see row [6] in Table C.1).

The baseline controls include cumulative real stock returns in R&D-intensive industries to capture any broad advanced information about future technological developments. Next, we add a broader set of financial indicators. Financial conditions could matter for several reasons, for instance, by determining the relative attractiveness of long-horizon investments in R&D, by summarizing additional forward-looking economic information with an influence on both productivity and government R&D, or more generally by capturing additional types of disturbances with potential effects on long-run productivity. We add the 3-month and 10-year Treasury rates, the log real S&P500 index, and the spread between BAA- and AAA-rated corporate bonds to the controls (obtained from FRED and Shiller (2015)). As can be seen from panel (a) in Figure C.3, these additional financial controls attenuate the TFP response somewhat at horizons beyond eight years. The TFP response at longer horizons remains highly statistically significant (see row [7] in Table C.1).

The next four specifications each, in turn, rotate in a number of additional variables that conceivably could contain important independent information about future productivity: Non-R&D capital in the business sector, the Fernald (2012) measure of labor quality, the patent-based innovation index of Cascaldi-Garcia and Vukotić (2022), and the relative price of R&D from the NIPA data. Including non-R&D capital leads to somewhat smaller estimates of the TFP response in the longer run, while including the relative price of R&D leads to estimates that are considerably larger. The addition of the indices for labor quality or innovation do not have any major impact on the estimates. As rows [8]-[11] in Table C.1 show, the estimates of the TFP response at longer horizons remain highly statistically significant in each case.

Table C.1: TFP Impact of Nondefense R&D Shock, Robustness

		% Impact After		
		5 years	10 years	15 years
[1]	Baseline	0.05 (−0.05,0.15)	0.18*** (0.09,0.27)	0.24*** (0.13,0.36)
[2]	+ Unemployment Rate	0.03 (−0.07,0.13)	0.20*** (0.08,0.32)	0.29*** (0.13,0.44)
[3]	+ Output Gap	0.05 (−0.06,0.17)	0.20*** (0.10,0.31)	0.28*** (0.13,0.42)
[4]	+ Defense R&D Appropriations	0.06 (−0.12,0.24)	0.22*** (0.06,0.38)	0.19** (0.00,0.37)
[5]	+ Public Infrastructure Capital	0.08* (−0.01,0.18)	0.15*** (0.06,0.25)	0.14*** (0.06,0.22)
[6]	+ Other Fiscal Variables	0.06 (−0.09,0.20)	0.07 (−0.05,0.19)	0.18*** (0.06,0.30)
[7]	+ Financial Variables	0.04 (−0.05,0.13)	0.11** (0.02,0.20)	0.18*** (0.09,0.27)
[8]	+ Non R&D Capital	0.04 (−0.06,0.14)	0.08** (0.01,0.15)	0.16*** (0.07,0.25)
[9]	+ Labor Quality	0.03 (−0.08,0.13)	0.16*** (0.09,0.24)	0.24*** (0.12,0.37)
[10]	+ Patent-Based Innovation Index	0.00 (−0.11,0.12)	0.18*** (0.06,0.30)	0.28*** (0.14,0.43)
[11]	+ Relative Price of R&D	−0.00 (−0.14,0.14)	0.24*** (0.09,0.39)	0.42*** (0.16,0.68)
[12]	Two Lags of Controls	0.07** (0.00,0.13)	0.16** (0.03,0.29)	0.16* (−0.01,0.34)
[13]	Six Lags of Controls	0.19 (−0.07,0.45)	0.45*** (0.22,0.67)	0.26*** (0.06,0.45)
[14]	Excluding Space	0.08 (−0.25,0.41)	0.20** (0.00,0.41)	0.25** (0.04,0.47)
[15]	Including Lags of Narrative Shock	0.07 (−0.05,0.19)	0.22*** (0.12,0.33)	0.24*** (0.16,0.32)
[16]	Balanced Sample	0.04 (−0.07,0.15)	0.17*** (0.08,0.25)	0.22*** (0.12,0.32)

Notes: Estimates are based on (2) using the narrative measure of federal nondefense R&D appropriations. Numbers in parentheses are 95 percent HAR confidence bands based on Lazarus et al. (2018). Stars $*$, $**$ and $***$ denote statistical significance at 10, 5, and 1 percent significance levels, respectively. Impulses scaled to imply a 1 percent peak increase in government R&D capital. Sample: 1948Q1–2021Q4 (specification with patent-based innovation index: 1949Q1–2010Q4).

C.4 Robustness: LP Model Specification

This section reports impulse response estimates of TFP to a nondefense R&D shock under several additional alterations to the baseline specification in (2). Panel (b) in Figure C.3 plots the impulse responses along with the baseline estimates and their 95 percent confidence bands from Figure 6 in the main text. Rows [12]-[15] in Table C.1 report the coefficient estimates for the various alterations at horizons of 5, 10, and 15 years with HAR confidence bands in parentheses.

The baseline specification uses $p = 4$ lags of all control variables. The first two robustness checks consider shortening or lengthening the number of lags to $p = 2$ and $p = 6$, respectively. As Panel (b) in Figure C.3 shows, reducing lag length from four to two quarters leads to somewhat smaller TFP responses at horizons beyond 10 years; the long-run TFP responses remain statistically significant at the 5 or 10 percent levels (see row [12] of Table C.1). Increasing the lag length from four to six quarters makes the TFP response somewhat more volatile, but the response at the end of the forecast horizon is very similar to the baseline specification and also remains highly significant (see row [13] of Table C.1).

As discussed in the main text, the rapid expansion of government R&D expenditures during the early stages of the space race is important for the precision of the estimates of the production function elasticities and rates of return reported in Tables 1 and 2. Our next robustness check analogously verifies the role of the early NASA R&D appropriations for the estimated TFP response to a nondefense R&D shock. We remove the influence of the early expansion during the space race by orthogonalizing the narrative measure of exogenous nondefense R&D shocks not only to the defense R&D measure, but also to all appropriations for NASA over the 1958–1963 period. Figure C.3b shows that the gradual rise in TFP following a nondefense R&D shock is robust to excluding the space race episode. Row [14] of Table C.1 shows that the long-run TFP response also remains significant at conventional levels, even though the confidence bands become notably wider.

The baseline set of controls includes four lags of the (log) of cumulative nondefense R&D appropriations, but not lags of the (orthogonalized) narrative R&D measures themselves. Figure C.3b shows that additionally including these lags has very little effect on the estimated TFP response and the associated confidence bands (see row [15] in Table C.1).

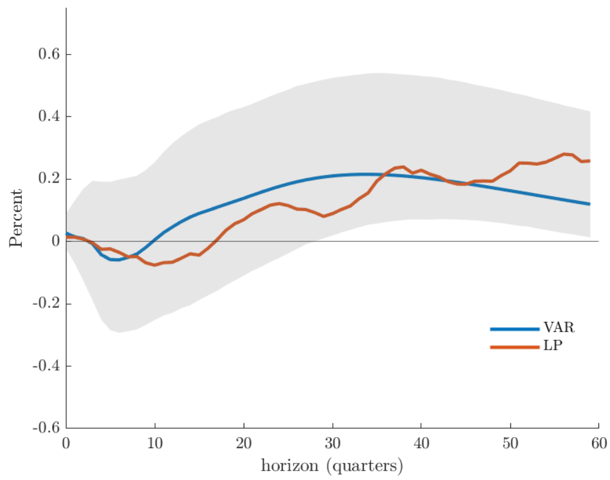
Finally, the inference formulas for SP-IV developed in Lewis and Mertens (2023) require a balanced sample. The impulse responses in Section 3 are instead estimated iteratively, i.e., using the largest possible estimation sample for each horizon h. Figure C.3b provides the estimated TFP response in the balanced sample, which leads to only relatively minor differences with the baseline estimates. As seen in row [16] of Table C.1, the estimates also remain highly statistically significant in the balanced sample.

C.5 Robustness: VAR-based Impulse Responses Estimates

Asymptotically, local projections estimate approximately the same impulse response as Vector Autoregressive Models (VARs) up to the lag length of the VAR model, see Plagborg-Møller and Wolf (2021). An advantage of LPs is that they avoid misspecification in finite-order VAR-based impulse response estimators at horizons beyond the lag length. In small samples, however, this advantage generally comes at the cost of greater variance, as shown for instance in the simulations of Li et al. (2022). In practice, VAR and LP impulse response estimates can differ meaningfully in small samples, raising questions about robustness.

Figure C.4 presents estimates of the TFP response to a nondefense R&D shock based on a VAR model, together with 95 percent confidence bands obtained using the wild bootstrap procedure described in Montiel Olea and Plagborg-Møller (2021). The estimates are

FIGURE C.4: TFP Impact of Nondefense R&D Shock, VAR Model Estimates



Notes: Estimates are based on an eight-variable VAR(4) model that includes all the variables from the baseline specification: the orthogonalized nondefense narrative measure, cumulative appropriations, (log) utilization-adjusted TFP, and the additional baseline controls described in the main text). VAR impulses are to an innovation in the narrative measure, scaled to imply a 1 percent peak increase in government R&D capital. The 95 percent confidence bands for the VAR impulse are percentile intervals based on the wild bootstrap described in Section 5 of Montiel Olea and Plagborg-Møller (2021). Sample: 1948Q1-2021Q4.

obtained from an 'internal instrument' VAR with four lags in eight variables: the orthogonalized nondefense narrative measure, (log) utilization-adjusted TFP, (log) cumulative sum of past changes in real nondefense R&D appropriations, and the additional controls of the baseline specification described in the main text. For comparison, the figure also shows the point estimates from the corresponding LP model.

As Figure C.4 shows, the VAR-based impulse response confirms our key finding: after a substantial delay, a positive shock to nondefense R&D appropriations leads to a gradual increase in business-sector TFP that becomes statistically significant in the long run. Overall, the magnitude of the VAR response is also similar to the LP response. The restrictions on the dynamics implied by the VAR do lead to some qualitative differences with the LP-based estimates. Specifically, the increase in TFP starts somewhat earlier and is hump-shaped. Despite these differences, we conclude that the positive long-run TFP response is robust to the choice of a VAR or LP-based impulse response estimator.

C.6 Robustness: Alternative LP Inference Procedures

The confidence intervals for the impulse responses are based on the equal-weighted cosine (EWC) test recommended by Lazarus et al. (2018). Herbst and Johanssen (2022) show in simulations that EWC delivers better empirical coverage than heteroskedasticity-and-autocorrelation robust (HAR) inference based on Newey and West (1987) or heteroskedastic-robust inference based on Ecker-Huber-White. Montiel Olea and Plagborg-Møller (2021) show that accounting for autocorrelation is redundant in lag-augmented LPs and that it suffices to use Ecker-Huber-White standard errors. The same authors also describe a wild bootstrap procedure that—in simulations of AR(1) models—delivers better coverage in small samples, especially at longer horizons and when the data is highly persistent.

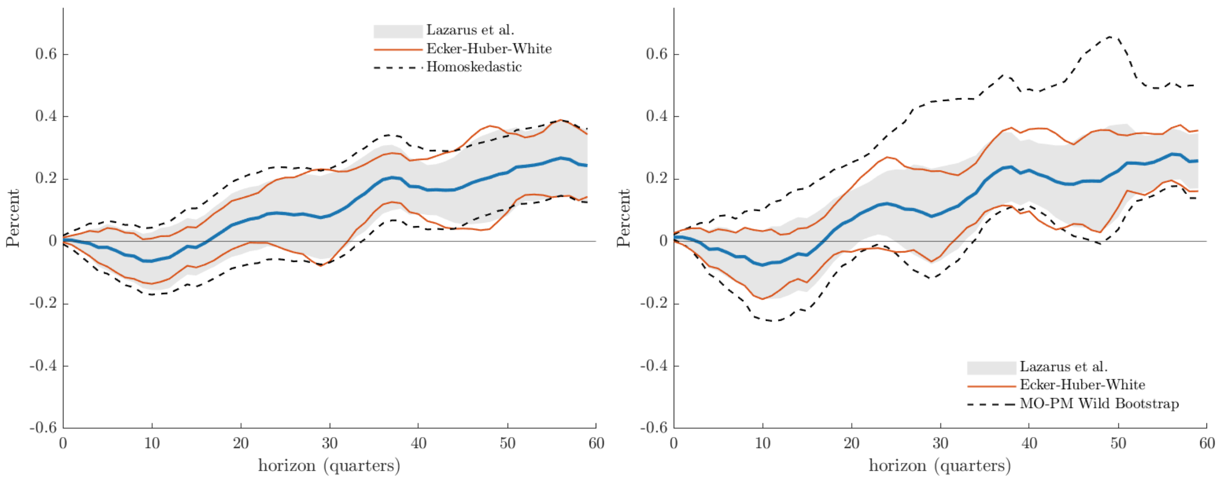
Figure C.5 compares various inference procedures for the impulse response of utilization-adjusted TFP based on the (orthogonalized) narrative measure for nondefense R&D appropriations. The left panel shows Ecker-Huber-White intervals and the simple intervals assuming homoscedasticity, along with the Lazarus et al. (2018) intervals, for the baseline specification with additional controls (same as in the bottom left panel of Figure 6). To capture a longer history of appropriations for R&D, the baseline specification includes lags of cumulative appropriations as controls rather than lags of the narrative measures. The right panel shows point estimates and confidence intervals based on specifications that additionally include four lags of the narrative measure, i.e., the explicit lag-augmented specification considered in Montiel Olea and Plagborg-Møller (2021). Apart from the Lazarus et al. (2018) intervals, the right panel again shows the Ecker-Huber-White intervals as well as the intervals based on the Montiel Olea and Plagborg-Møller (2021) wild bootstrap procedure.

The main conclusion from Figure C.5 is that the choice of inference procedures is relatively unimportant. The homoscedastic and Ecker-Huber-White bands are similar to the Lazarus et al. (2018) EWC bands. The wild bootstrap bands are meaningfully wider, but the increase in coverage lies mostly to the north of the Lazarus et al. (2018) region. Especially at longer horizons, the lower bootstrap band is relatively close to the Lazarus et al. (2018) band. Importantly, the finding that a shock to nondefense appropriations leads to a statistically significant long-run increase in business-sector TFP is not affected by the choice of inference procedures.

C.7 Impact of a Defense R&D Shock on Other Productivity/Innovation Indicators

Figure 7 in the main text reports the impact of a nondefense R&D shock on various productivity measures and innovation indicators. Figure C.6 reports the impact of a defense R&D shock on the same variables. Whereas a positive nondefense R&D shock consistently leads to increases in all productivity and innovation indicators, the same is not the case for defense R&D shocks. Figure C.6 shows a hump-shaped transitory decline in labor productivity and no statistically or economically significant impact on potential output. There also are transitory declines in the patent innovation index and the number of Ph.D. recipients in STEM fields. The number of R&D researchers increases in the short run, but declines in the longer run. There is no meaningful change in the number of technology publications, except perhaps at longer horizons.

Figure C.5: TFP Impact of Nondefense R&D Shock, Alternative Inference Procedures



Notes: All confidence intervals are for the 95 percent level. *Left Panel*: Point estimates and shaded confidence intervals (Lazarus et al. (2018) HAR) are identical to those in the bottom left panel of Figure 6 (baseline specification). *Right Panel*: Point estimates and shaded confidence intervals (Lazarus et al. (2018) HAR) are based on the baseline specification with four lags of the nondefense narrative measure added to the controls. The figure also shows bootstrap intervals as described in Section 5 of Montiel Olea and Plagborg-Møller (2021), based on 10,000 samples. Impulses are scaled to imply a 1 percent peak increase in government R&D capital. Sample: 1948Q1-2021Q4.

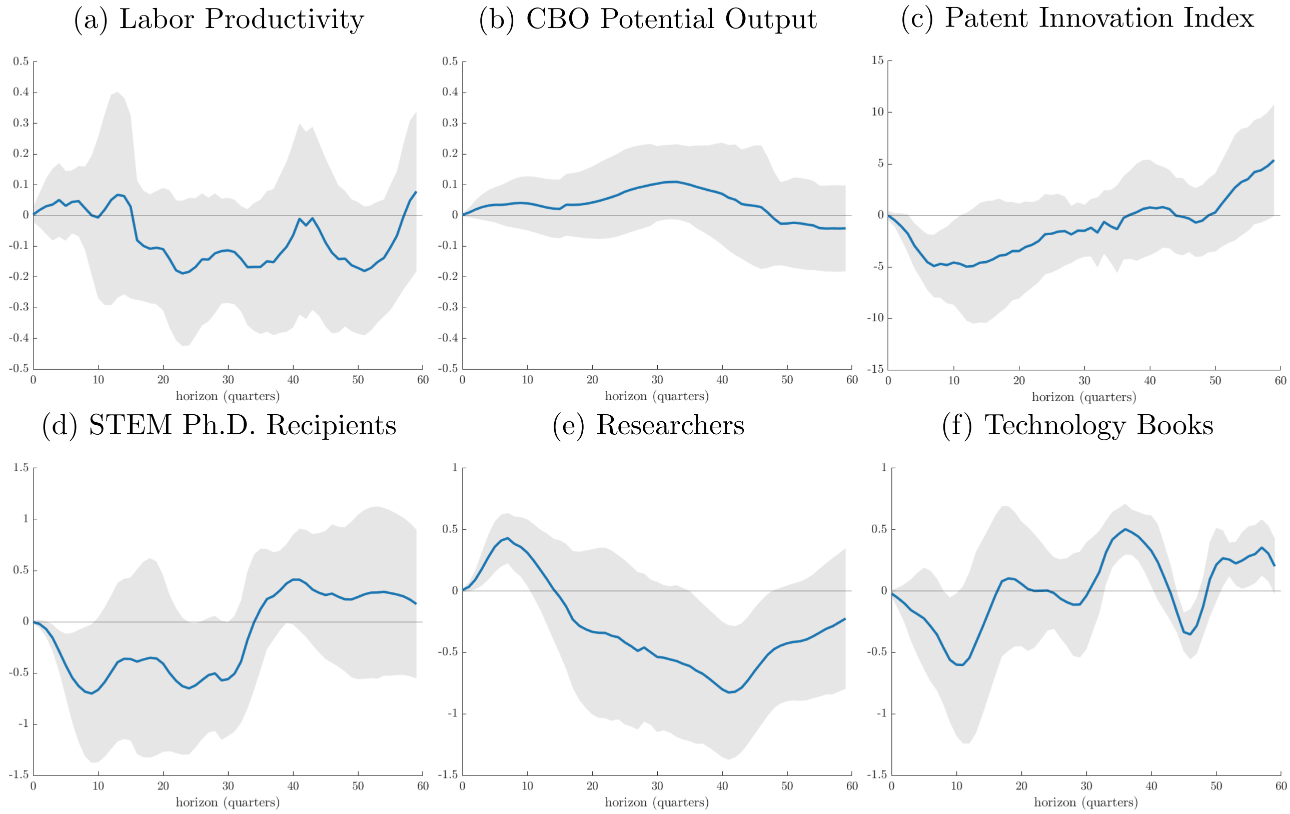
C.8 Responses of Private Labor and non-R&D Capital Inputs

Figure C.7 shows estimates of the responses of other private factor inputs following positive shocks to nondefense (panel a) and defense (panel b) R&D appropriations. The measures of private factor inputs are from Fernald (2012). The estimates are obtained from local projections as in (2) in the main text, with the same baseline controls and four lags of each outcome variable added as additional controls. As in Figures 6 and 7 in the main text, the impulse responses are scaled to imply a one percent peak increase in the total government R&D capital stock. The first row in Figure C.7 depicts responses of labor input adjusted for labor quality (cumulative sum of '`dhours`' + '`dLQ`' in F-TFP, see Appendix A). The second row shows the responses of the business-sector non-R&D capital stock, which consists of all types of capital excluding R&D and software (nonresidential equipment and structures, residential business structures, and non-R&D intellectual property).

The first row in Figure C.7 shows that a nondefense R&D shock leads to little change in (quality-adjusted) labor input in the business sector at most horizons. Towards the end of the 15-year forecast horizon, there is a decline in labor input that is marginally statistically significant at one or two horizons. The response of labor input to a defense R&D shock is somewhat volatile and imprecisely estimated, with none of the estimates statistically significantly different from zero at the 5 percent level.

The second row in Figure C.7 shows that, with a long delay, a nondefense shock leads to a gradual and persistent increase in the business-sector non-R&D capital stock that is highly

FIGURE C.6: Impact of a Defense R&D Shock on Other Productivity/Innovation Indicators

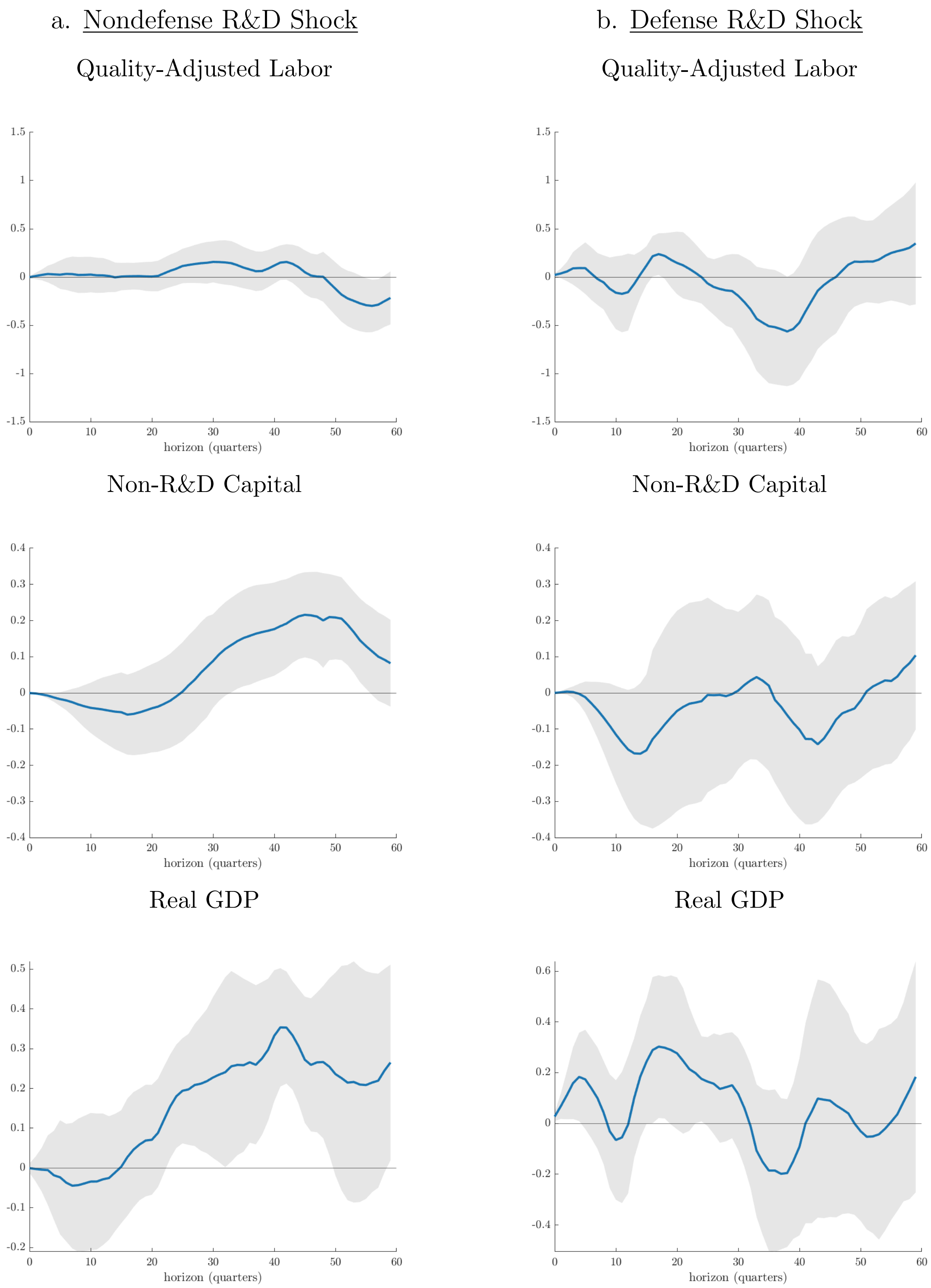


Notes: Estimates based on (2) using the orthogonalized narrative measure of changes in defense R&D appropriations, see (1). Lazarus et al. (2018) HAR bands are for 95 percent confidence levels. Impulses scaled to imply a 1 percent peak increase in government R&D capital. Sample: (a),(b),(d): 1948Q1–2021Q4; (c): 1949Q1–2010Q4; (e): 1951Q1–2019Q4; (f): 1956Q1–1997Q4. See Appendix A for variable definitions.

statistically significant at horizons between 6 to 14 years. The peak increase in non-R&D capital is roughly 0.2 percent and occurs after about 13 years. The response of non-R&D capital to a defense R&D shocks shows some evidence of a transitory decline in the short run but is overall imprecisely estimated.

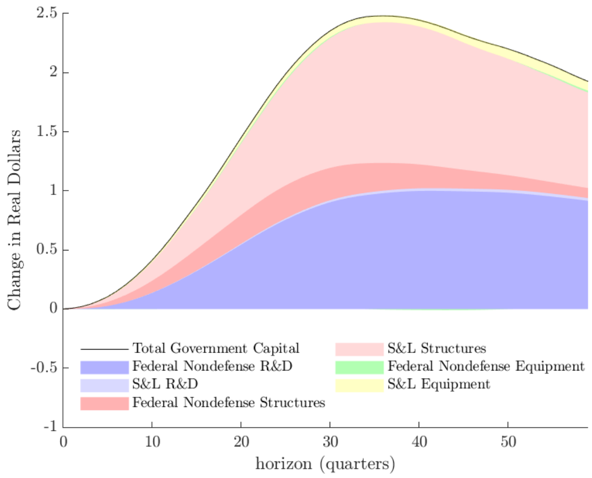
The final row in Figure C.7 shows the responses of real GDP. A nondefense shock does not lead to any economically or statistically significant change in real GDP in the short run. In the longer run, real GDP increases by around 0.2 to 0.35 percent. The timing and magnitude of the GDP response are overall similar to that of business-sector labor productivity or potential output, see Figure 7 in the main text. The response of real GDP to a defense R&D shock is positive and marginally significant at a few horizons over the first five years, but the point estimates are imprecisely estimated at longer horizons and oscillate between positive and negative. Consistent with the impulses to defense shocks shown in Figures 6 and C.7, there is no evidence that defense shocks lead to a significant long-run increase in real GDP.

Figure C.7: Labor and non-R&D Capital Following an Increase in R&D Appropriations



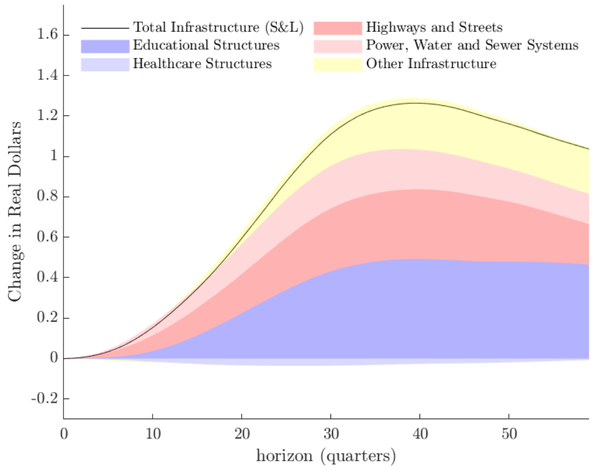
Notes: Estimates based on (2) using the orthogonalized narrative measure of changes in federal nondefense (left panel) and defense (right panel) R&D appropriations, see (1). 'Baseline' includes additional lagged controls described in the main text. Lazarus et al. (2018) HAR bands are for 95 percent confidence levels. Impulses scaled to imply a 1 percent peak increase in government R&D capital. Sample: 1948Q1–2021Q4.

FIGURE C.8: Nondefense Public Capital



Notes: Estimates based on (2) using the orthogonalized narrative measure of changes in federal nondefense R&D appropriations, see (1). Impulses are scaled to imply a unit peak increase in federal nondefense R&D capital. Sample: 1948Q2–2021Q4.

FIGURE C.9: S&L Structures by Function



Notes: Estimates based on (2) using the orthogonalized narrative measure of changes in federal nondefense R&D appropriations, see (1). Impulses are scaled to imply a peak increase in state and local structures of 1.21 dollars, to match Figure C.8. Sample: 1948Q2–2021Q4.

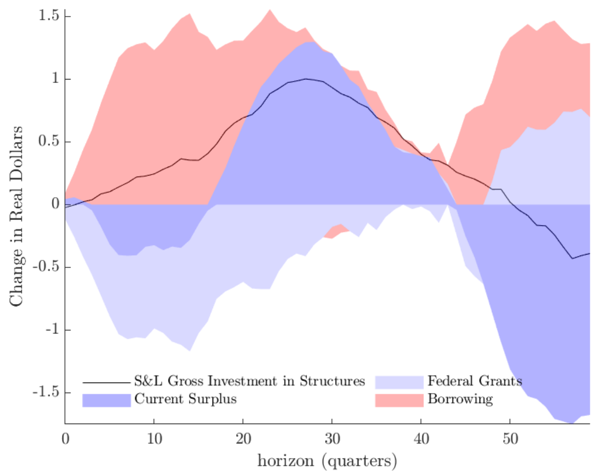
C.9 A Closer Look at the Public Infrastructure Response to a Nondefense Shock

Figure 9 in the main text shows that an increase in appropriations for nondefense R&D leads to a rise in public infrastructure, and specifically in nondefense structures. In this section, we present further decompositions similar to those in Figure 9 to better understand the nature of the rise in public infrastructure after a nondefense R&D shock.

The first additional decomposition considers the response of various components of total nondefense public capital by type and level of government, i.e., federal versus state and local (S&L) government. Figure C.8 shows that the increase in public infrastructure after a nondefense shock is primarily driven by a rise in structures funded by state and local governments (up to 1.19 dollars), although there is also an increase in federal infrastructure spending on structures (up to 28 cents). Note that the total increase does not exactly add up to the 1.58 dollar increase seen in Figure 9 because of slight differences in the regression specifications (the lagged outcome variables y_{t-j} on the right-hand side in (2) are different). The main text, therefore, reports the contribution of state and local government structures as a percentage $(1.19/(1.19 + 0.28) \approx 0.8)$.

Figure C.9 provides a further breakdown of the state and local government infrastructure response into various categories based on additional detail in the BEA Fixed Assets Accounts (Table 7.1), with quarterly values obtained by interpolation of the annual source data. The responses, in this case, are scaled to match the peak 1.21 dollar increase in Figure C.8. As the figure shows, the largest increase occurs in educational structures. There are also meaningful increases in highways and streets as well as in power, water, and sewer systems.

Figure C.10: Financing of S&L Investment in Structures



Notes: Estimates based on (2) using the orthogonalized narrative measure of changes in federal nondefense R&D appropriations, see (1). Impulses are scaled to imply a unit peak increase in S&L gross investment in structures. Sample: 1949Q1–2021Q4.

The changes in all remaining types of state and local government infrastructure ('Other Infrastructure') are individually relatively small.

Figure C.10 provides a breakdown of the response of investment in structures by state and local governments according to the means of financing: Debt, federal transfers, or current surpluses (revenues less other spending). Note that, unlike in the previous figures, this decomposition pertains to the flow (real gross investment in structures) rather than the stock (the capitalized real cost value of structures). The decomposition is based on the budget constraint identity aggregated across state and local governments using data from the BEA (NIPA Table 3.3). The impulses are scaled to imply a unit peak increase in S&L gross investment in structures.

Figure C.10 shows that, consistent with the response of the corresponding capital stock, a nondefense R&D shock leads to a gradual rise in state and local investment in nondefense structures. Investment peaks after about seven years, subsequently returns to prior levels, and towards the end of the forecast horizon, even dips slightly below the level predicted in the absence of the nondefense R&D shock. Figure C.10 also shows that the investment boom is not financed by increased federal transfers to state and local governments. The latter initially fall and only revert to prior levels well after the peak in investment. For the first couple of years, the rise in investment is accounted for by an increase in borrowing by state and local governments. Between horizons of 4 to 10 years, the investment boom is implicitly financed by a surplus in revenues relative to other state and local spending. The main takeaway from Figure C.10 is that the rise in state and local investment in nondefense structures does not appear to be driven by increases in federal grants to state and local governments, for instance, to increase spending on highways.

D Estimation of Production Function Elasticity: Additional Results

This section presents additional results for the estimation of the production function elasticity of government R&D capital ϕ in Section 4 in the main text.

D.1 SP-IV as a Regression in Impulse Response Space

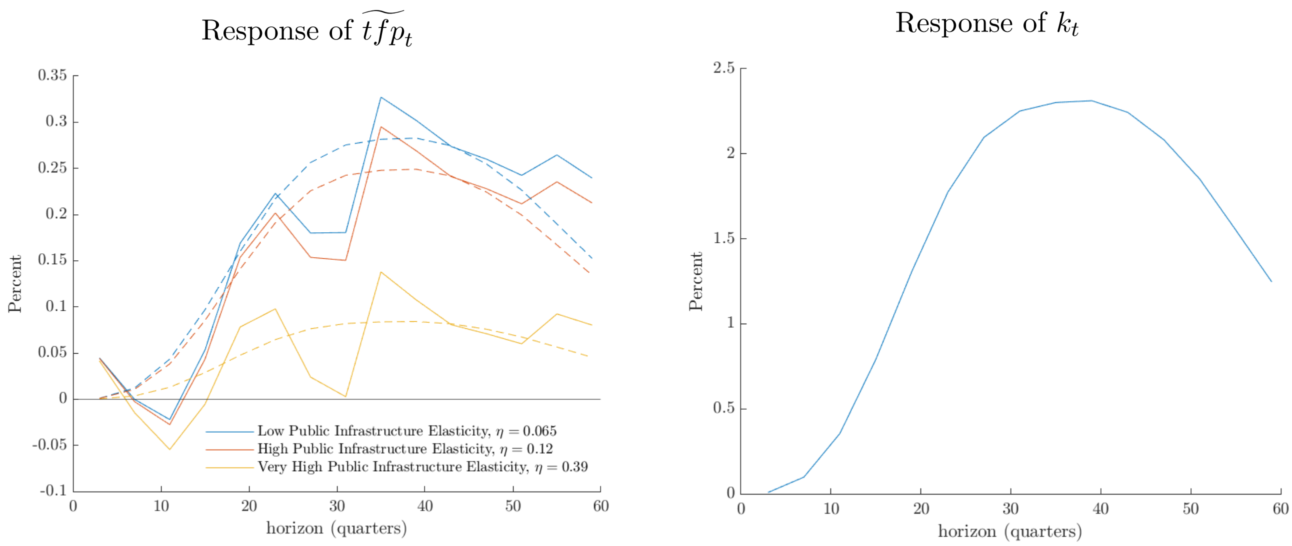
Figure D.1 provides the main intuition behind the SP-IV estimation of ϕ in (7) based on the response to the orthogonalized narrative measure of nondefense R&D appropriations, z_t^{ND}, using the specification in (2). The solid lines in the left panel show the response of $\widetilde{tfp}_t$ to a one standard deviation innovation in z_t^{ND} for three different values of η, and the right panel shows the estimated response of k_t, the government R&D capital stock. Both figures show the impulse responses at one-year intervals that used for the estimation of the production function elasticity. The left panel shows the response for the endpoints of Ramey's (2021) plausible range, $\eta = 0.065$ and $\eta = 0.12$; to make the dependence on η visually clearer, the figure also shows the response for a much higher value $\eta = 0.39$, which is the estimate in Aschauer (1989). The SP-IV estimate of ϕ in each case is simply the OLS coefficient $\hat{\phi}$ in a regression (without intercept) of the impulse response coefficients of $\widetilde{tfp}_t$ in the left panel on those of k_t in the right panel. The dashed lines in the left panel show the resulting fitted values—$\hat{\phi}$ times the impulse response of k_t—that minimize the sum of squared residuals for each value of η. The SP-IV regression framework thus estimates the structural parameter as the value of ϕ that best fits the relationship between $\widetilde{tfp}_t$ and k_t along the impulse response trajectories. The functional form in (7) imposes very specific assumptions on the lags between R&D spending and the TFP effects. As Figure D.1 shows, the dynamics of the fitted TFP responses align well with those of the actual TFP responses, such that the timing assumptions implied by the structural equation appear reasonable in light of the responses estimated in the local projections.

SP-IV can make use of more than one set of impulse response coefficients for identification, e.g., to both defense and nondefense shocks, in which case the inverse covariance matrix of the identifying innovations weights the different impulse responses. The SP-IV estimator also applies to structural equations with multiple endogenous regressors, as in specification (9) in the main text, in which case it reduces to multiple regression in impulse response space, see Lewis and Mertens (2023).

D.2 Simultaneous Confidence Sets

For the specifications with two endogenous regressors, i.e., (9) and (11) in the main text, the confidence intervals reported in Tables 1 and 2 are subvector confidence sets obtained using

Figure D.1: Illustration of the SP-IV Estimator



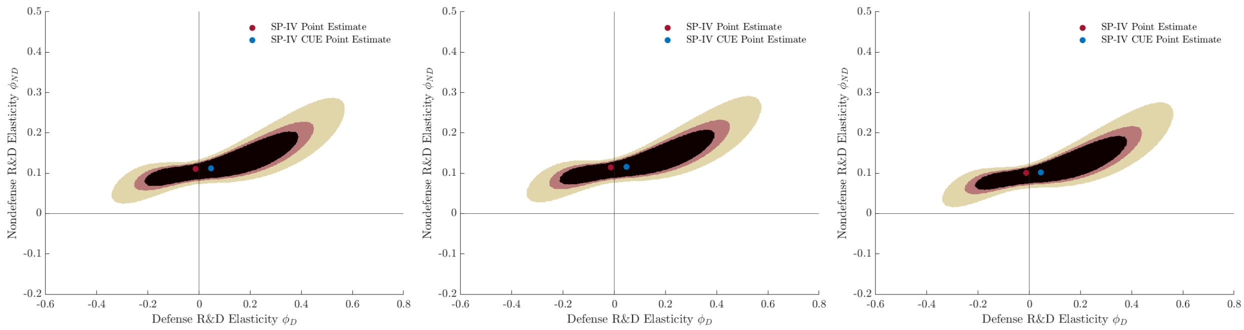
Notes: Solid lines show impulse response estimates (at one-year intervals) to a one standard deviation innovation in the orthogonalized narrative measure of changes in nondefense R&D appropriations using the baseline specification in (2) in a balanced sample. The SP-IV estimator $\hat{\phi}$ results from regressing the impulse response coefficients of $\widetilde{tfp}_t$ in the left panel on the impulse response coefficients of k_t in the right panel without intercept, see Lewis and Mertens (2023). The dashed lines in the left panel show the fitted responses obtained by multiplying $\hat{\phi}$ by the response of k_t in the right panel.

the projection method, see, e.g., Andrews et al. (2019). As an illustration, the panels in Figure D.2 show the 68, 90, and 95 percent weak-instrument-robust confidence sets for the full parameter vector $[\phi_{ND}, \phi_D]$ associated with the estimates reported in row [6] of Table 1. The confidence intervals reported in Table 1 for $\hat{\phi}_{ND}$ ($\hat{\phi}_D$) are the largest and smallest values of $\hat{\phi}_{ND}$ ($\hat{\phi}_{ND}$) across all values of ϕ_D ($\hat{\phi}_{ND}$) that belong to the 95 percent simultaneous confidence set. The simultaneous confidence sets are based on inverting the KLM statistic of Kleibergen (2005). The latter is based on the score of the continuously updated Anderson-Rubin statistic (or equivalently, the S-statistic of Stock and Wright (2000) for GMM) as a function of ϕ_{ND} and ϕ_D, see Lewis and Mertens (2023). The minimum of the Anderson-Rubin objective does not correspond to the SP-IV point estimate, such that the latter does not generally lie at the 'center' (or is even within) of the confidence sets. An alternative estimator of (ϕ_{ND}, ϕ_D) is the minimand of the continuously-updated Anderson-Rubin objective function, which by construction lies at the 'center' of the confidence sets. This continuously updated estimator (CUE) is marked by the blue dots in Figure D.2. As can be seen from the figure, the CUE estimates of ϕ_{ND} are all very close to the SP-IV estimates, whereas those for ϕ_D are marginally larger.

Figure D.3 shows the simultaneous confidence sets for the three remaining specifications in Table 1 that include nondefense and defense capital separately (rows [7]-[9]). For brevity, the figure reports only the confidence sets for the specifications that assume the interme-

Figure D.2: Simultaneous Weak-Instrument Robust Confidence Sets

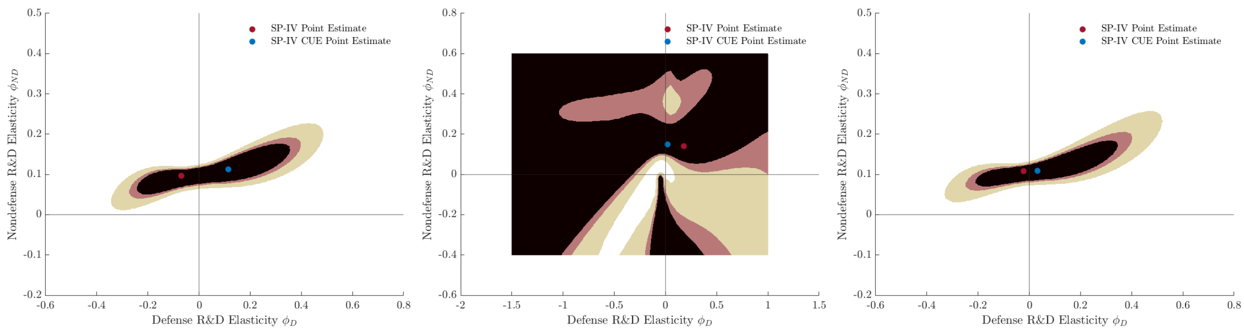
(a) Intermediate $\eta = 0.08$ (b) Low $\eta = 0.065$ (c) High $\eta = 0.12$



Notes: Confidence sets based on inverting the KLM statistic of Kleibergen (2005) for the specification in row [6] of Table 1.

Figure D.3: Simultaneous Weak-Instrument Robust Confidence Sets

(a) Using Both Shocks (b) Excluding Space (c) All Appropriations



Notes: Confidence sets based on inverting the KLM statistic of Kleibergen (2005) for the specification in rows [7]-[9] of Table 1 for $\eta = 0.08$.

diate value of the infrastructure elasticity, $\eta = 0.08$. As can be seen from the figures, the CUE estimate is usually close to the SP-IV estimate, and always nearly identical for the nondefense elasticity. The simultaneous confidence sets are also very similar across specifications. The exception is the specification with the narrative measure that excludes the large appropriations for the space race, see panel (b) in Figure D.3. For that specification, the confidence sets have highly irregular shapes, and most values of either parameter cannot be ruled at conventional levels of confidence.

D.3 Wald Inference

In the main text, inference for the SP-IV estimates is based on the weak-instrument robust methods for GMM described in Kleibergen (2005). Lewis and Mertens (2023) show that the SP-IV estimator is equivalent to a restricted 2SLS estimator in a system of equations,

Table D.1: SP-IV Elasticity Estimates with Wald Inference

	Public R&D Measure	Instruments	Intermediate η ϕ_{ND}	Intermediate η ϕ/ϕ_D	Low η ϕ/ϕ_{ND}	High η ϕ/ϕ_{ND}
[1]	Total	Exo ND	0.12*** (0.08,0.16)		0.12*** (0.08,0.17)	0.11*** (0.06,0.15)
[2]	Total	Exo ND, No Space	0.14*** (0.05,0.24)		0.14*** (0.05,0.24)	0.13*** (0.04,0.22)
[3]	Total	All ND	0.11*** (0.07,0.16)		0.12*** (0.08,0.16)	0.10*** (0.06,0.14)
[4]	Total	Exo D		−0.24 (−0.69,0.20)		
[5]	Total	All D		−0.23 (−0.67,0.21)		
[6]	ND/D	Exo ND	0.11*** (0.06,0.16)	−0.01 (−0.33,0.30)	0.11*** (0.07,0.16)	0.10*** (0.05,0.15)
[7]	ND/D	Exo ND/D	0.10*** (0.05,0.14)	−0.07 (−0.35,0.21)	0.10*** (0.06,0.14)	0.09*** (0.05,0.13)
[8]	ND/D	Exo ND, No Space	0.14*** (0.04,0.24)	0.18 (−0.75,1.10)	0.14*** (0.04,0.25)	0.13** (0.03,0.23)
[9]	ND/D	All ND	0.11*** (0.07,0.15)	−0.02 (−0.33,0.28)	0.11*** (0.07,0.15)	0.10*** (0.06,0.14)

Notes: See notes to Table 1 in the main text. The only difference is that the confidence intervals are based on the Wald formulas derived under the assumption of strong identification, see Lewis and Mertens (2023).

where the number of equations is equal to the number of impulse response horizons used for identification. Under strong identification and otherwise standard assumptions, this formulation of the SP-IV estimator leads to conventional Wald inference formulas. It is well known that—when identification is weak—Wald inference can suffer from large size distortions in small samples, and the simulations in Lewis and Mertens (2023) show that this is also the case for the SP-IV estimator. Table D.1 shows the same point estimates as Table 1 in the main text, but reports confidence intervals based on the conventional Wald formulas. Qualitatively, the only specification for which there are large differences in the inference results is the one in row [8], i.e., the specification with the narrative measure that excludes the large appropriations for the space race: The Wald-based inference points to estimates that are highly statistically significant, whereas the weak-instrument-robust inference result leads to the conclusion that the instrument is uninformative. The estimates of the defense R&D capital elasticity, on the other hand, remain insignificant also under Wald inference.

D.4 Specification with Constant Elasticities

In specification (9) in the main text, the production function elasticities of defense and nondefense R&D capital scale with their nominal shares in total government R&D capital.

TABLE D.2: GOVERNMENT R&D PRODUCTION FUNCTION ELASTICITIES
ALTERNATIVE SPECIFICATION

	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]	ND/D	Exo ND	0.07^{**} (0.02,0.13)	0.16 (−0.42,0.47)	0.07^{**} (0.02,0.13)	0.06^{**} (0.01,0.12)
[2]	ND/D	Exo ND/D	0.08^{**} (0.01,0.12)	−0.03 (−0.30,0.37)	0.08^{**} (0.01,0.13)	0.08^{*} (−0.00,0.12)
[3]	ND/D	Exo ND, No Space	0.13 (−2.00,0.11)	−0.09 (−0.93,2.00)	0.13 (−2.00,0.11)	0.12 (−2.00,0.10)
[4]	ND/D	All ND	0.07^{***} (0.02,0.13)	0.13 (−0.41,0.43)	0.07^{***} (0.02,0.13)	0.06^{**} (0.01,0.12)

Notes: Rows [1]-[4] show SP-IV estimates of ϕ_{ND} (nondefense) and ϕ_D (defense) in (D.1). All specifications include the baseline set of lagged controls described in Section 3. Numbers in parentheses are weak-instrument robust confidence intervals at the 5 percent significance level based on inverting the KLM statistic of Kleibergen (2005). Test inversion is limited to a grid with endpoints −2 and 2, † denotes intervals constrained at these endpoints. Subvector inference is based on the projection method. $*$, $**$ and $***$ denote statistical significance at 10, 5 and 1 percent levels respectively. 'Exo ND/D' denotes the orthogonalized narrative measure of exogenous changes in nondefense/defense R&D appropriations. 'All ND' denotes the orthogonalized series of all changes in nondefense/defense R&D appropriations, ignoring our narrative classification. 'No Space' indicates that the instrument is also orthogonalized to all changes in space appropriations between 1958 and 1963.

The following specification instead imposes constant elasticities:

$$\Delta \widetilde{tfp}_t = \phi_{ND}\left(\bar{s}_{ND}\Delta k_t^{ND}\right) + \phi_D(1-\bar{s}_{ND})\Delta k_t^D + \Delta w_t \tag{D.1}$$

We multiply the regressors by the average shares, $\bar{s}_{ND}$ and $1-\bar{s}_{ND}$, over the estimation sample, such that the estimates are on a comparable scale to those reported in Table 1 in the main text. The estimation results based on (D.1) are reported in Table D.2. The estimates can be multiplied by $\bar{s}_{ND} \approx 0.5$ to obtain the elasticities with respect to Δk_t^{ND} and Δk_t^D.

The main difference with the results in the main text is that the point estimates for ϕ_{ND} are smaller. The only exception is in row [3], but this is also the specification for which the estimates are very imprecise. Ignoring the results in row [3], the point estimates of ϕ_{ND} are around 0.07, as compared to 0.12 under the specification discussed in the main text. The estimates of ϕ_{ND} are relatively precisely estimated (except in row 3]), and they are highly statistically significant. Just as in the main text, the estimates of ϕ_D vary considerably across the specifications. They are always imprecise and never statistically distinguishable from zero.

The difference in the estimates of ϕ_{ND} between the specification in equation (9) and the one in (D.1) is not too surprising, given that the share of nondefense R&D varies considerably over the estimation sample. Given that the stock of nondefense R&D capital is small in the beginning of the sample, the log differences Δk_t^{ND} are very large early on, which leads to lower overall estimates of ϕ_{ND}. Weighting by the shares as in the baseline specification (9) in the main text attenuates the influence of these early observations, and should therefore lead to more accurate estimates for the whole sample.

Even if one would prefer the lower estimates in Table D.2, they do not change the overall conclusion that the rate of return on nondefense government R&D is very high. Dividing the estimates in rows [1], [2], and [4] of Table D.2 by 0.06 (the average ratio of government R&D capital to GDP), the implied rates of returns range from 100 to 150 percent.

Finally, note that the point estimates of ϕ_{ND} in row [3] of Table D.2 lie outside of the reported weak-instrument-robust confidence intervals. As explained in Appendix D.2, this is possible with the confidence sets based on Kleibergen (2005) as they are not necessarily centered on the GMM estimates.

D.5 Different Depreciation Rates

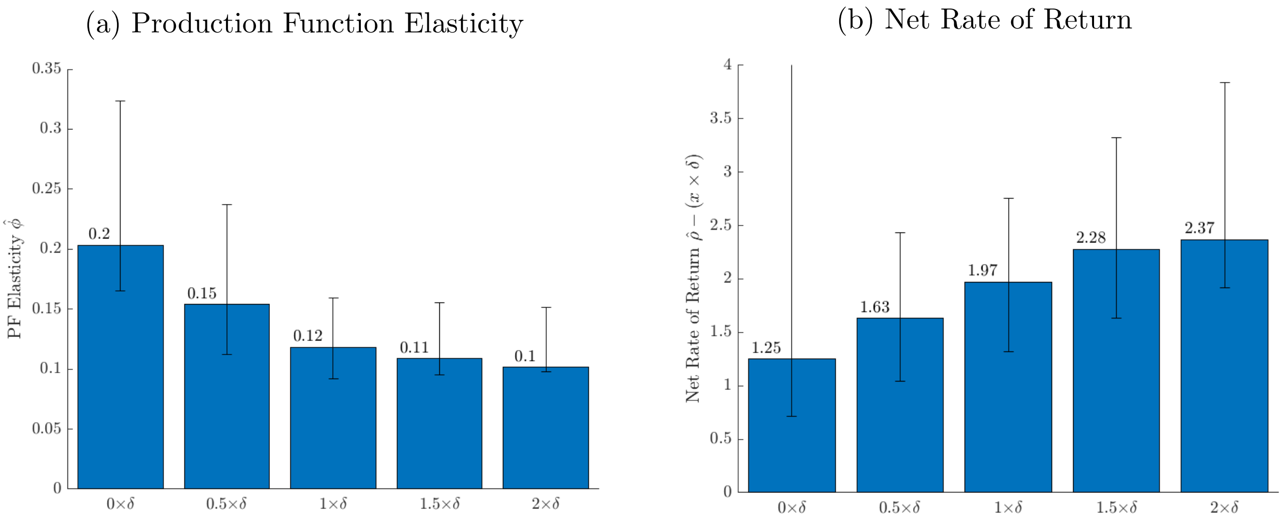
The quarterly measures of the government R&D capital stocks that we use throughout the analysis closely follow the methodology of the BEA, which publishes the annual totals as part of the 'Fixed Assets' tables. For certain categories of government R&D, the BEA estimates depreciation rates based on observing a progression of specific R&D investments with observable outcomes on the effective life of the R&D. For other categories of government R&D, the BEA uses the same depreciation rates as for private R&D services.

Given the inherent difficulties in measuring the obsolescence of intellectual capital, we verify how the estimates of the production function elasticities and rates of return change under different assumptions about depreciation rates on government R&D. Specifically, we capitalize the various categories of government R&D investment by multiplying the annual BEA depreciation rates for each category by a scaling factor $x = 0, 0.5, 1, 1.5$ or 2. On average across (weighted) categories and years, the BEA depreciation rate is $\delta \approx 0.16$. When $x = 0$, all depreciation rates are zero. When $x = 2$, all depreciation rates are twice as large as those used by the BEA, therefore averaging to $2 \times \delta \approx 0.32$. For simplicity, we keep the initial values of each subcomponent of the R&D capital stock constant to the 1946 values in the BEA tables.

Figure D.4 shows how the estimation results (all assuming $\eta = 0.08$) change with the assumed depreciation rates. The left panel shows the estimates of the production function elasticity, obtained exactly as in row [1] of Table 1. The right panel shows the estimates of the net rate of return, obtained by estimating the gross rate of return exactly as in row [1] of Table 2 and subtracting the (scaled) average depreciation rate. The error bars mark the 95 percent weak-instrument-robust confidence intervals.

As the left panel of Figure D.4 shows, the production function elasticity estimates are decreasing in the assumed depreciation rate. Intuitively, assuming a larger depreciation rate implies a smaller estimate of the net stock of R&D capital, and therefore, a one percent increase in the capital stock corresponds to a smaller increase in investment expenditures.

Figure D.4: Nondefense Government R&D,
Elasticity and Return Estimates Assuming Different Depreciation Rates



Notes: SP-IV estimates of ϕ (left) and rates of return ρ (right) based on (7) and (10), respectively. Estimates are based on the (orthogonalized) narrative measure of nondefense appropriations as in rows [1] of Table 1 and 2, respectively, and assuming the intermediate value $\eta = 0.08$. Error bars are 95 percent weak-instrument-robust confidence intervals based on inverting the KLM statistic of Kleibergen (2005).

As mentioned in the main text, the BEA depreciation rates result in elasticity estimates that are centered around 0.12. Assuming zero depreciation raises the point estimate of the elasticity to 0.20, whereas doubling the depreciation rates lowers the estimate to 0.10. The right panel of Figure D.4 shows that the net rate of return is increasing in the assumed depreciation rate. Although the elasticity estimates are decreasing in the depreciation rate, larger depreciation rates also lower the capital stock to GDP ratio estimate, which translates to higher rates of return. Using the BEA estimates, the point estimate of the net rate of return is $(2.13 - 0.16) \times 100 = 197$ percent. This estimate drops to 125 percent, assuming zero depreciation. Doubling the depreciation rates increases the net return estimate to 237 percent. Even if one would prefer to assume a higher or lower average depreciation rate on intellectual capital, doing so would not change the main conclusion that the rate of return on nondefense government R&D is relatively high.

Online Appendix References

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

Andrews, Isaiah, James H. Stock, and Liyang Sun (2019). "Weak Instruments in Instrumental Variables Regression: Theory and Practice". In: *Annual Review of Economics* 11.1, pp. 727–753.

Antolin-Diaz, Juan and Paolo Surico (2022). *The Long-Run Effects of Government Spending*. Discussion Paper 15062. Centre for Economic Policy Research.

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

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.

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

Cloyne, James, Joseba Martinez, Haroon Mumtaz, and Paolo Surico (2022). *Short-Term Tax Cuts, Long-Term Stimulus.* Working Paper 30246. National Bureau of Economic Research.

Croce, M.M., Thien T. Nguyen, S. Raymond, and L. Schmid (2019). "Government debt and the returns to innovation". In: *Journal of Financial Economics* 132.3, pp. 205–225.

Fernald, John G. (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". mimeo.

Herbst, Edward P. and Benjamin K. Johanssen (2022). *Bias in Local Projections.* mimeo. Board of Governors.

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.

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

Li, Dake, Mikkel Plagborg-Møller, and Christian Wolf (2022). *Local Projections vs. Vars: Lessons from Thousands of Dgps.* NBER Working Paper 30207. Princeton University.

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.

Montiel Olea, José Luis and Mikkel Plagborg-Møller (2021). "Local Projection Inference Is Simpler and More Robust Than You Think". In: *Econometrica* 89.4, pp. 1789–1823.

Newey, Whitney K. and Kenneth D. West (1987). "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix". In: *Econometrica* 55.3, pp. 703–708.

Plagborg-Møller, Mikkel and Christian K. Wolf (2021). "Local Projections and VARs Estimate the Same Impulse Responses". In: *Econometrica* 89.2, pp. 955–980.

Ramey, Valerie A. (2021). “The Macroeconomic Consequences of Infrastructure Investment”. In: *Economic Analysis and Infrastructure Investmentn.* 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.

Shiller, Robert J. (2015). *Irrational Exuberance, 3rd ed.* Princeton, NJ: Princeton University Press.

Stock, James H. and Jonathan H. Wright (2000). “GMM with Weak Identification”. In: *Econometrica* 68.5, pp. 1055–1096.

Whelan, Karl (2002). “A Guide To U.S. Chain Aggregated Nipa Data”. In: *Review of Income and Wealth* 48.2, pp. 217–233.