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.

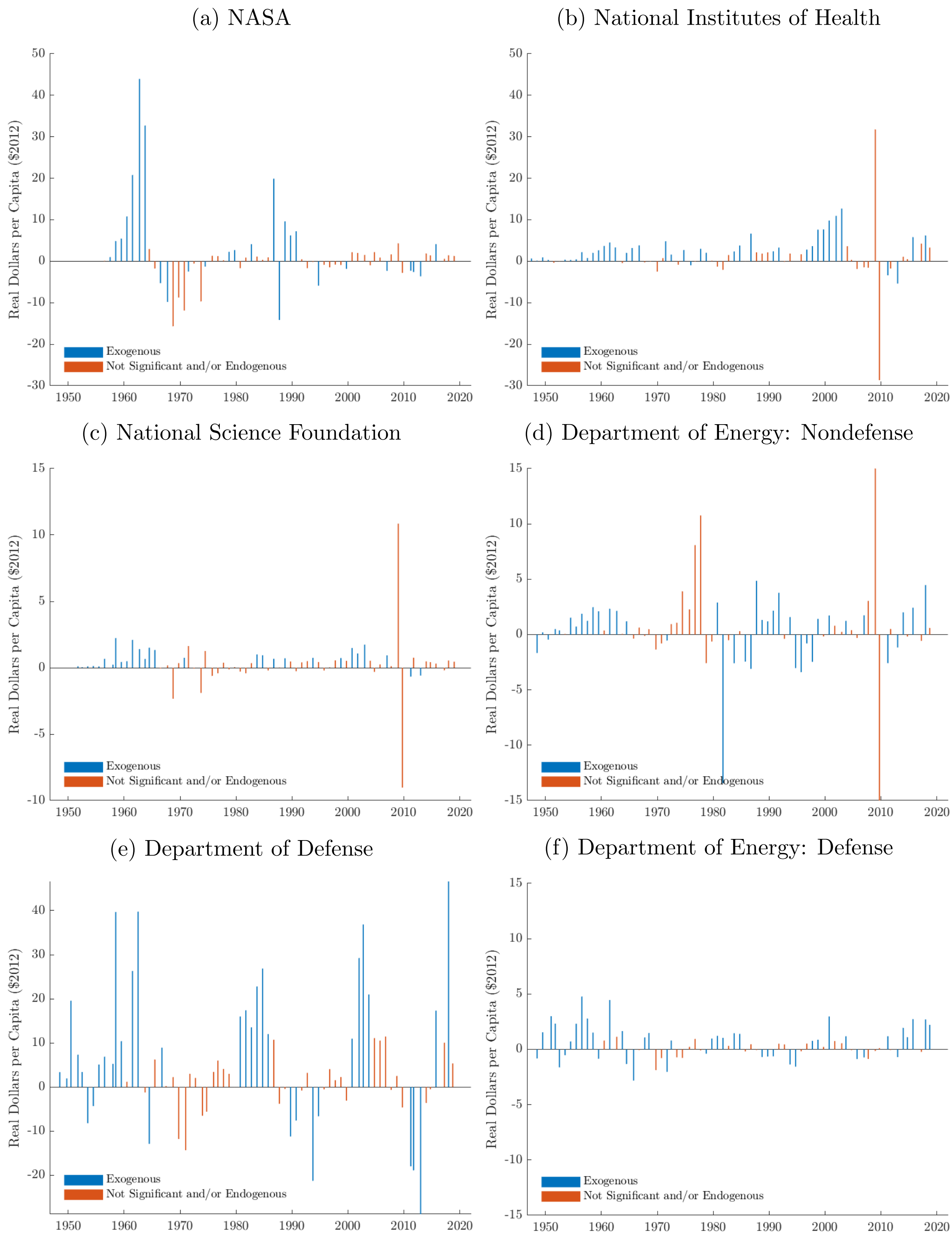
Other variables:

- 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),

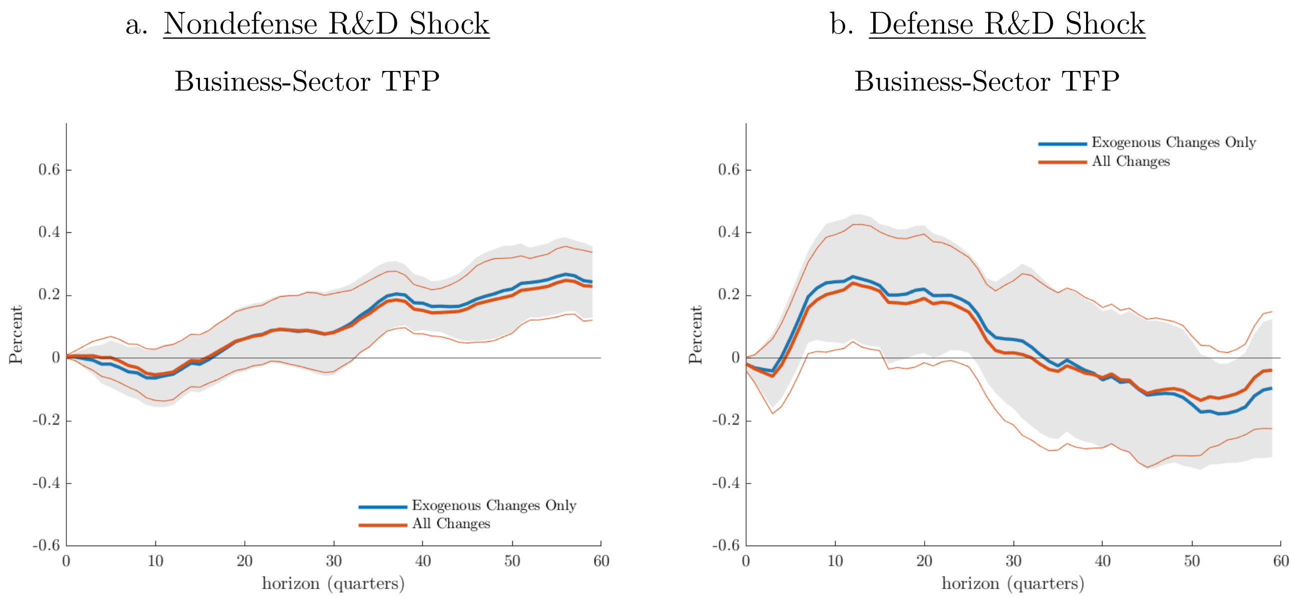
FIGURE B.1: Changes in R&D Appropriations by Federal Agency



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.

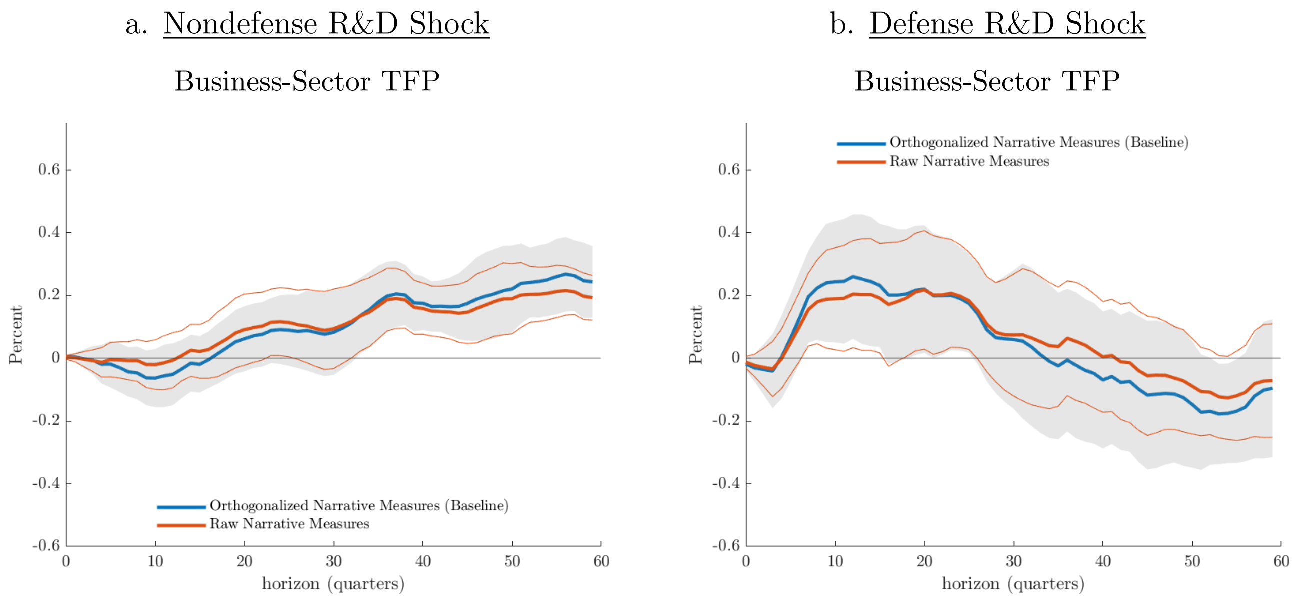
R&D appropriations shocks for defense expenditures, with DOD in panel (e) and the nuclear security functions of DOE in panel (f). Appropriations shocks classified as exogenous are depicted in blue, and those classified as endogenous (or too small to classify) are in red; all R&D appropriations shown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

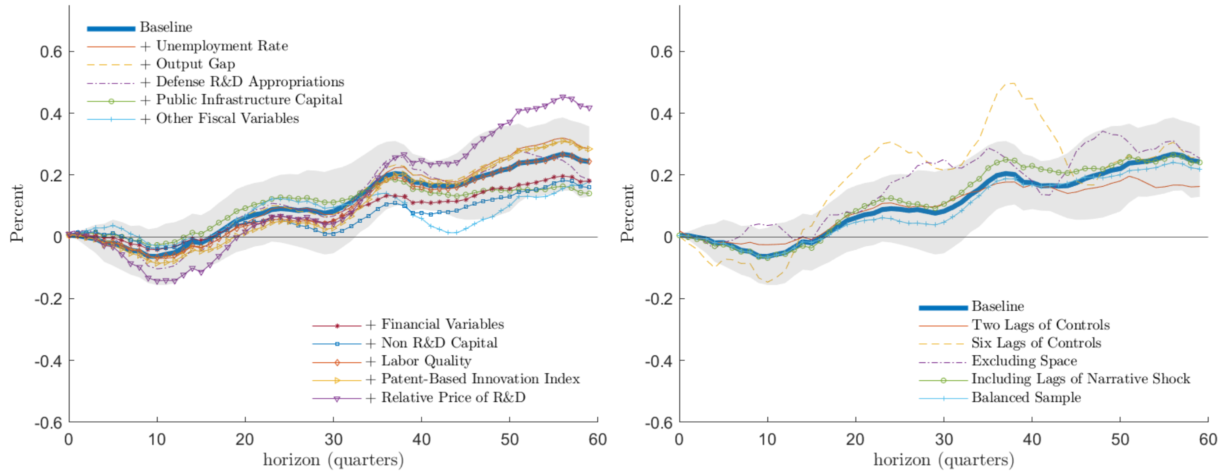
This section discusses the role of mutually orthogonalizing the narrative measures of exogenous changes in defense and nondefense R&D appropriations for the impulse response estimates, as in equation (1) in the main text. Figure C.2 replicates the baseline impulse responses of TFP to nondefense and defense 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).

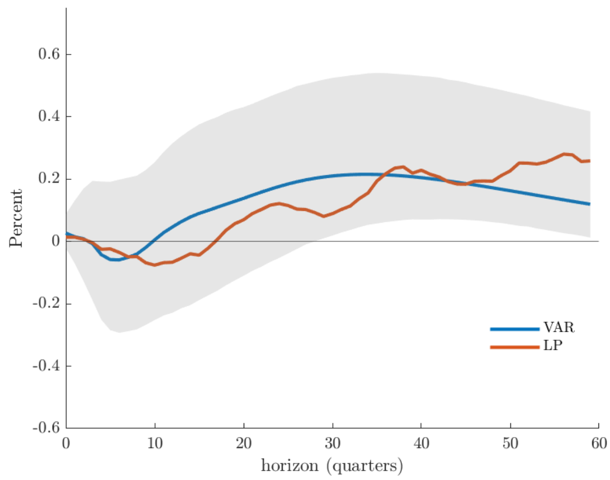
Finally, the inference formulas for SP-IV developed in Lewis and Mertens (2023) require a balanced sample. The impulse responses in Section 3 are instead estimated iteratively, i.e., using the largest possible estimation sample for each horizon h. Figure C.3b provides the estimated TFP response in the 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.

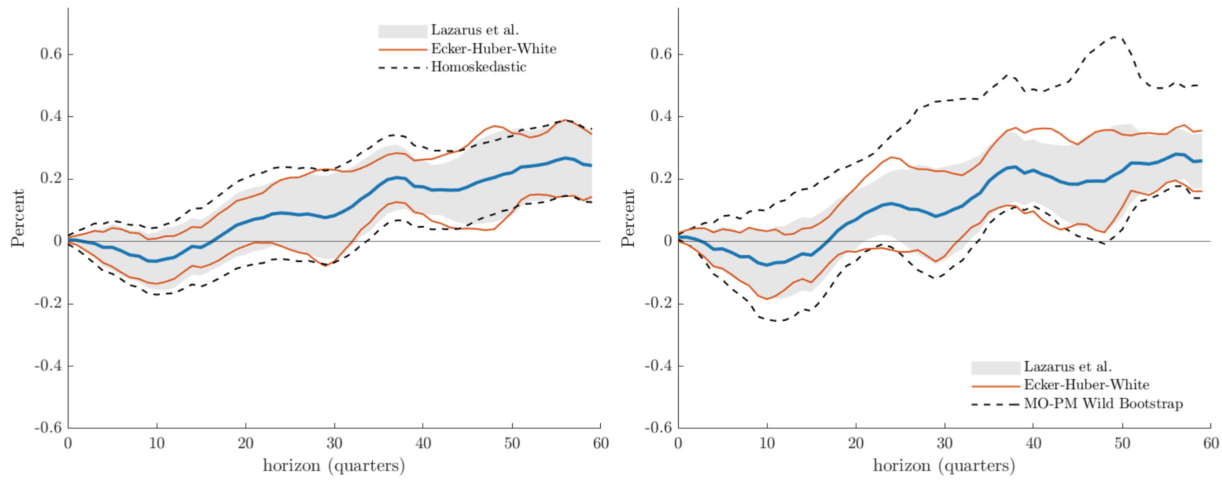
Figure C.5 compares various inference procedures for the impulse response of utilization-adjusted TFP based on the (orthogonalized) narrative measure for nondefense R&D appropriations. The left panel shows Ecker-Huber-White intervals and the simple intervals assuming homoscedasticity, along with the 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.

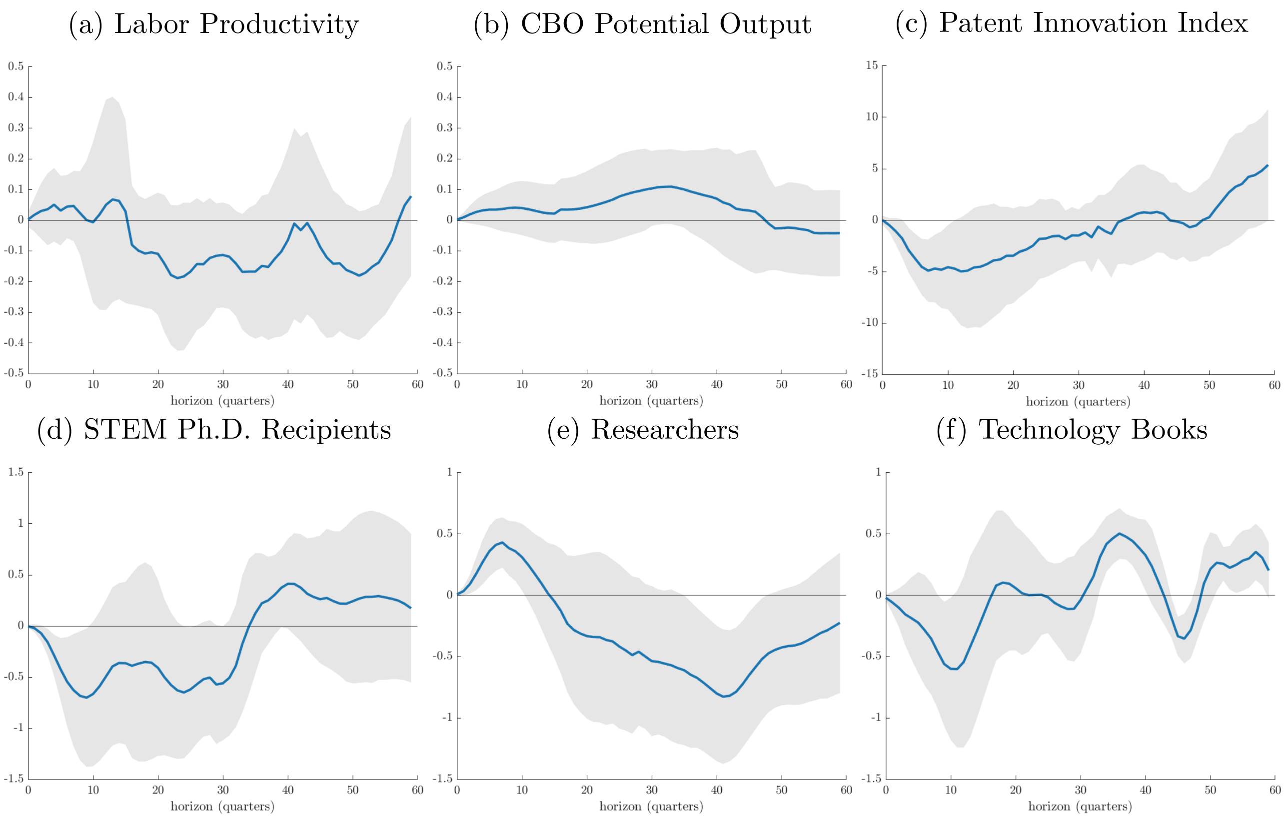
C.8 Responses of Private Labor and non-R&D Capital Inputs

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

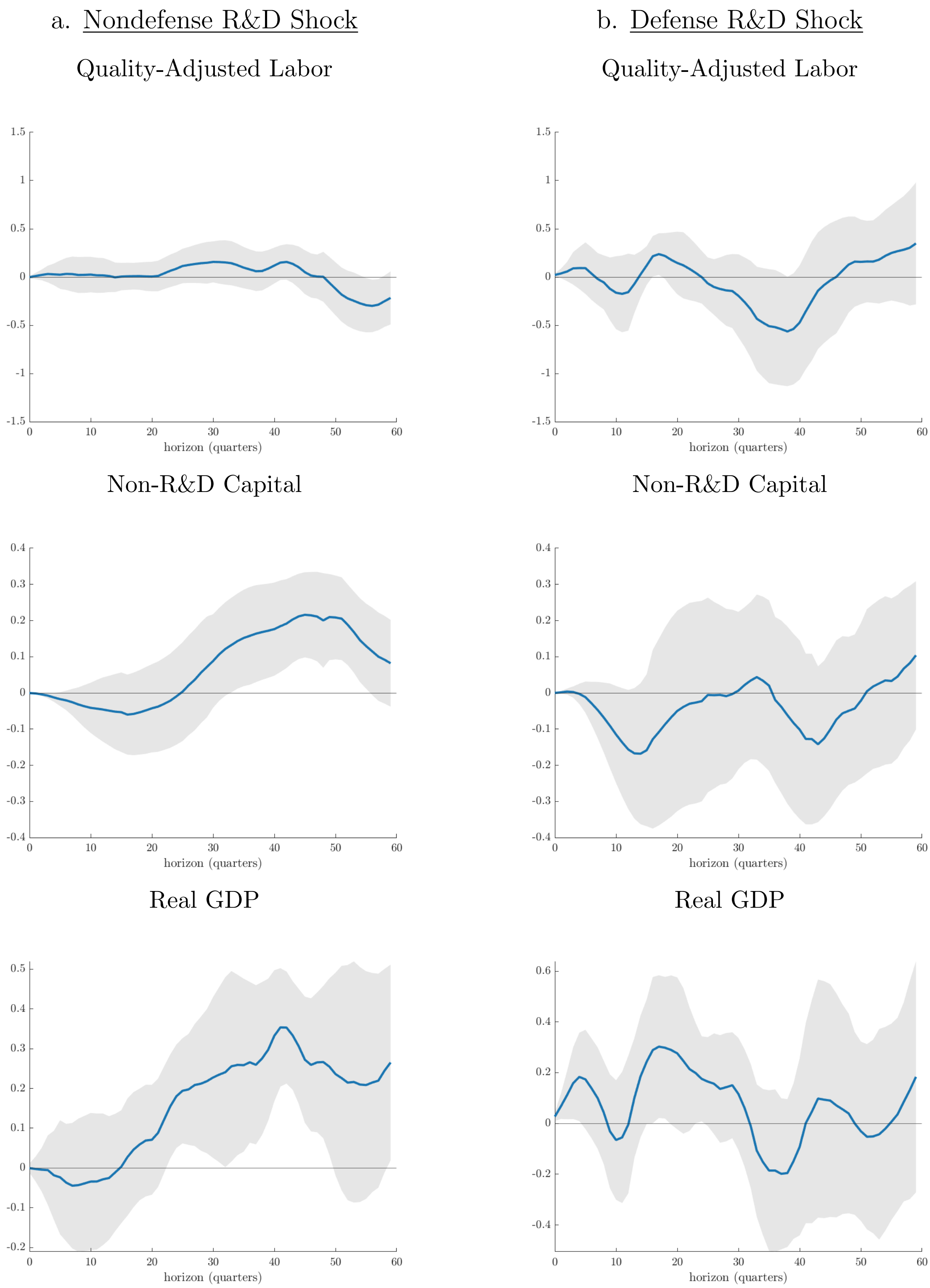


Notes: Estimates based on (2) using the orthogonalized narrative measure of changes in defense R&D appropriations, see (1). Lazarus et al. (2018) HAR bands are for 95 percent confidence levels. Impulses scaled to imply a 1 percent peak increase in government R&D capital. Sample: (a),(b),(d): 1948Q1–2021Q4; (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.

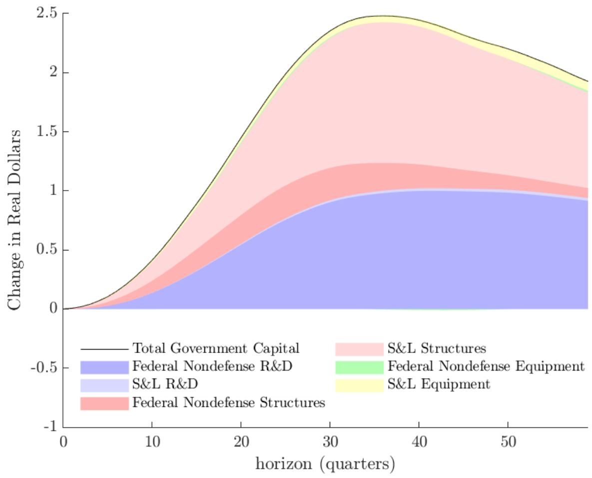
The final row in Figure C.7 shows the responses of real GDP. A nondefense shock does not lead to any economically or statistically significant change in real GDP in the short run. In the longer run, real GDP increases by around 0.2 to 0.35 percent. The timing and magnitude of the GDP response are overall 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



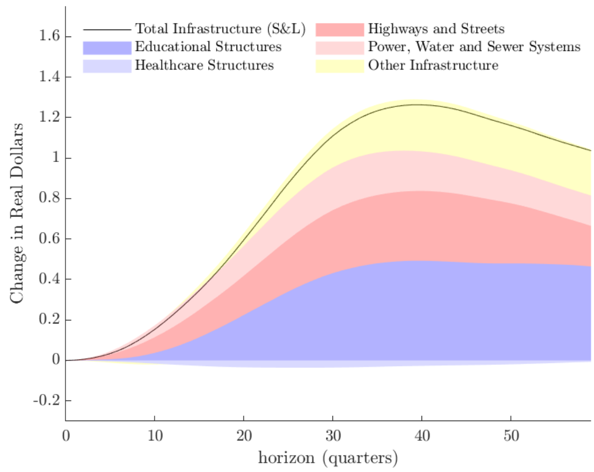
Notes: Estimates based on (2) using the orthogonalized narrative measure of changes in federal nondefense (left panel) and defense (right panel) R&D appropriations, see (1). 'Baseline' includes additional lagged controls described in the main text. Lazarus et al. (2018) HAR bands are for 95 percent confidence 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.

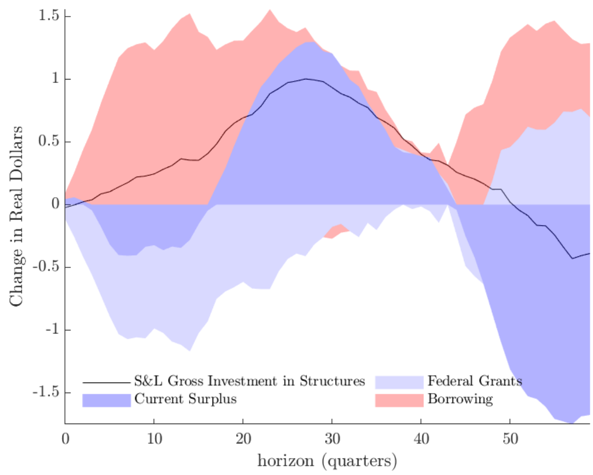
C.9 A Closer Look at the Public Infrastructure Response to a Nondefense Shock

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

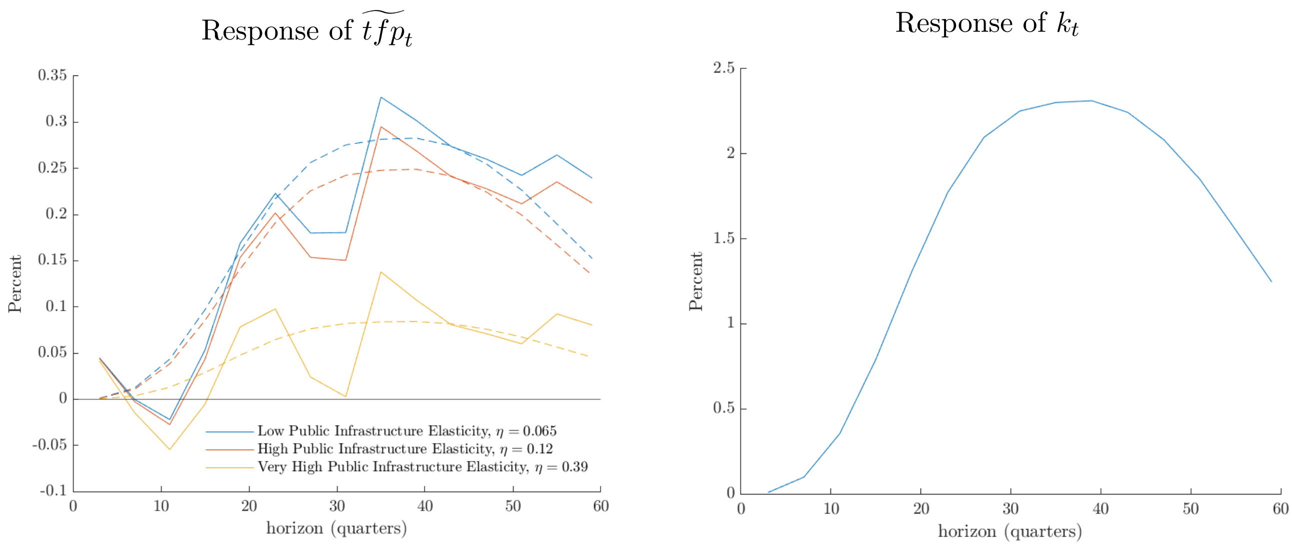
Figure D.1 provides the main intuition behind the SP-IV estimation of ϕ in (7) based on the response to the orthogonalized narrative measure of nondefense R&D appropriations, z_t^{ND}, using the specification in (2). The solid lines in the left panel show the response of $\widetilde{tfp}_t$ to a one 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



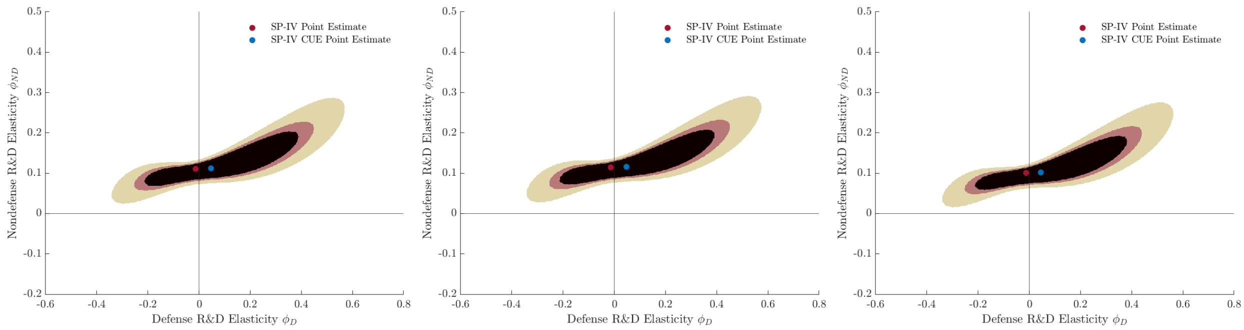
Notes: Solid lines show impulse response estimates (at one-year intervals) to a one standard deviation innovation in the orthogonalized narrative measure of changes in nondefense R&D appropriations using the baseline specification in (2) in a balanced sample. The SP-IV estimator $\hat{\phi}$ results 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

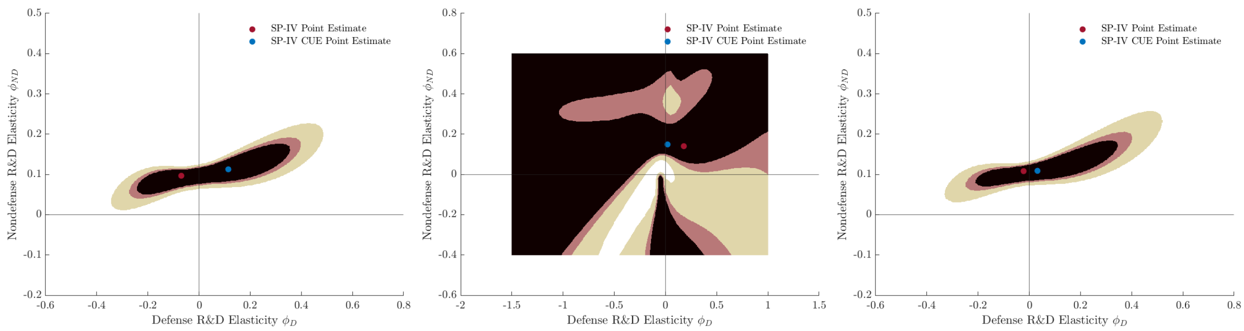
(a) Intermediate $\eta = 0.08$ (b) Low $\eta = 0.065$ (c) High $\eta = 0.12$



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:

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

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

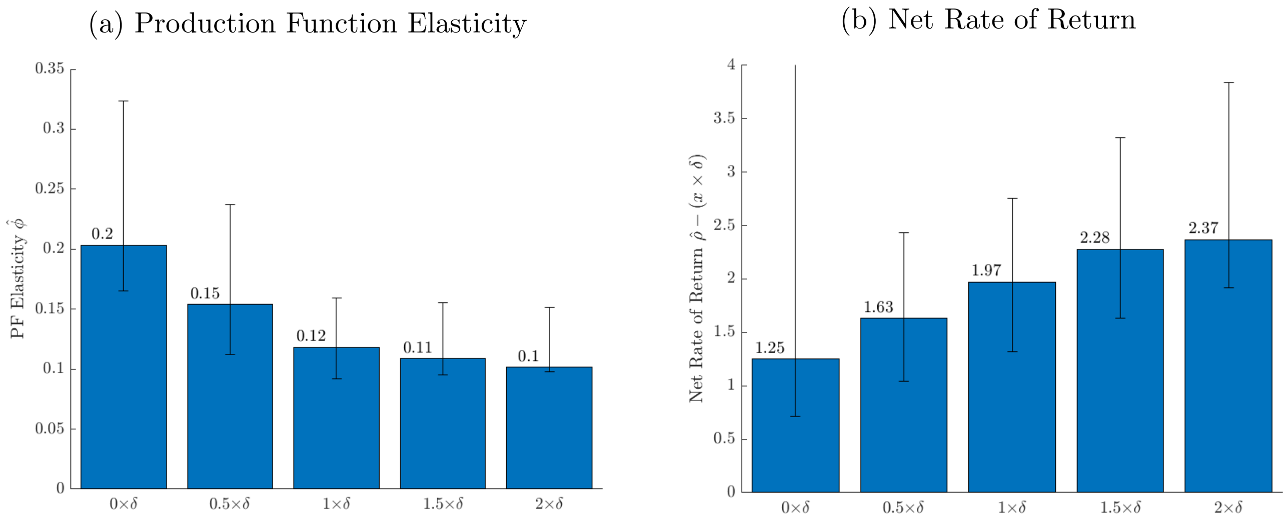
The quarterly measures of the government R&D capital stocks that we use throughout the analysis closely follow the methodology of the BEA, which publishes the annual totals as part of the 'Fixed Assets' tables. For certain categories of government R&D, the BEA estimates depreciation rates based on observing 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.