American Economic Review 2025, 115(7): 2376–2413
https://doi.org/10.1257/aer.20231278

The Long-Run Effects of Government Spending†

By Juan Antolin-Diaz and Paolo Surico*

Military spending has large and persistent effects on output because it shifts the composition of public spending toward R&D. This boosts innovation and private investment in the medium term and increases productivity and GDP at longer horizons. Public R&D expenditure stimulates economic activities beyond the business cycle even when it is not associated with war spending. In contrast, the effects of public investment are shorter-lived, while public consumption has a modest impact at most horizons. We reach these conclusions using BVAR with long lags and 125 years of US data, including newly reconstructed series of government spending by main categories since 1890. (*JEL* E21, E22, E23, E62, H50, H56, O30)

Can government spending stimulate long-run growth? Large increases in public expenditure—typically associated with defense buildups around wars—have often been credited with the development of new technologies. For instance, the Manhattan Project during WWII led to the development of nuclear energy, the establishment of the Defense Advanced Research Projects Agency (DARPA) in the late 1950s is linked to the creation of the Internet, and NASA's moon landing program of the 1960s spurred several advances in aeronautics and satellite technology, such as GPS. Despite this anecdotal evidence, the macroeconomics literature has not yet established a causal link between large government programs and long-term productivity, innovation, and growth at the aggregate level.

Using the series of military spending news constructed by Ramey and Zubairy (2018) (which builds on Ramey and Shapiro 1998; Ramey 2011b), we find that the effects of an unanticipated increase in defense spending are large and extend well beyond the frequencies typically studied in business cycle analyses. The output multiplier (i.e., the dollar increase in GDP that results from a dollar increase in government spending) is around one in the short run but rises significantly above

* Antolin-Diaz: London Business School (email: jantolindiaz@london.edu). Surico: London Business School and CEPR (email: psurico@london.edu). Emi Nakamura was the coeditor for this article. We are grateful to the coeditor, David Abraham, Michele Andreolli, Isaiah Andrews, Tim Besley, Olivier Blanchard, Gabriel Chodorow-Reich, Thomas Drechsel, Luca Gambetti, Will Hotten, Ethan Ilzetzki, Diego Känzig, Elias Papaioannou, Ivan Petrella, Mikkel Plagborg-Møller, Franck Portier, Valerie Ramey, Giovanni Ricco, Juan Rubio-Ramirez, Andrew Scott, Anna Rogantini Picco, Jon Steinsson, Christian Wolf, and seminar participants at the Barcelona Summer Forum 2022, the Monetary Economics workshop of the NBER Summer Institute 2022, Rutgers University, Sveriges Riksbank, and Norges Bank for very useful comments and suggestions. Many thanks to four anonymous referees for very insightful advice. Paolo Surico gratefully acknowledges financial support from the ESRC (Grant ref. ES/Y002490/1). The figures of this paper are best appreciated on screen or when printed in color.

† Go to https://doi.org/10.1257/aer.20231278 to visit the article page for additional materials and author disclosure statement(s).

one at long horizons. Total factor productivity (TFP), innovation, consumption, and investment fall in the first years after the shock but then recover and increase persistently over the medium term.[1]

As for the transmission mechanism, we present evidence that military spending stimulates the economy over the medium term because it shifts the composition of public spending toward R&D. While in the short run government consumption, investment in equipment and structures, and R&D all increase following military spending, public R&D is the only category that still responds significantly 10 years after the shock. To tease out the different effects across components, we use an alternative strategy that identifies the shock that maximizes the variance of each spending category within the first year after the shock. We find that persistent increases in output and TFP are associated with shocks that expand the share of government spending going to R&D.

Finally, we scrutinize our newly identified "public R&D shock" and show not only that it is weakly correlated with war spending but also that its historical evolution aligns well with narrative evidence on large R&D federal spending programs, including the Manhattan Project, DARPA, the moon landing program, and Reagan's "Star Wars" initiative. Furthermore, we document that an exogenous increase in public R&D leads to very sustained responses in output, TFP, innovation, and investment that are even larger and more persistent than the effects of military spending on these variables. Our results highlight a new channel through which fiscal policy can support economic activities in peacetime.

Identifying very persistent dynamics requires long, high-quality historical data and empirical methods suited to capture low-frequency correlations. As for the historical data, we have digitized archival statistics and drawn upon narrative evidence to construct new quarterly series of US government spending since 1890, by main categories: consumption expenditure, equipment and structures investment, and R&D. We have also constructed quarterly series for aggregate hours worked, total factor productivity, private investment and consumption, export and imports, building on existing and unpublished annual and quarterly data. This allows us to examine the effects of government spending at any relevant frequency over a period of 125 years that spans major military conflicts and public spending programs, financial crises and recessions, and monetary policy and fiscal policy regimes.

As for the empirical method, we rely on Bayesian Vector Autoregressions (BVAR) with a very long lag structure to compute dynamic causal effects. This approach allows us to capture the gradual patterns of technological diffusion after increases in R&D. It also connects us with the debate in empirical macro about the relative merits between VARs and direct single-equation regressions, known as "local projections" (LPs) (Jordà 2005; Kilian and Lütkepohl 2017; Nakamura and Steinsson 2018). Recent work has highlighted the intimate connection between the two approaches, and in particular, their coincidence up to the lag-order of the VAR (Plagborg-Møller and Wolf 2021). Moreover, Li, Plagborg-Møller, and Wolf (2021b) highlight the nontrivial bias-variance trade-off inherent to the choice between methods, and the attractiveness of shrinkage estimators in this context. We

[1] Throughout the paper, we will use interchangeably "low frequency," "long lasting," "long run," "medium term," "intermediate," and "long horizons" to refer to persistent dynamics that extend beyond business cycle frequencies.

set the lag order of the BVAR equal to 60 quarters, our maximum horizon of interest in the impulse responses, and employ shrinkage to maximize the marginal likelihood of the model (as in Giannone, Lenza, and Primiceri 2015), balancing these statistical considerations. We show that our main findings of significant effects of fiscal policy on output and productivity beyond business cycle frequencies are a robust feature of the US data that emerge also when we (i) exclude WWII or any other cluster of military events from the instrument, (ii) employ alternative model and prior specifications, or (iii) use alternative econometric methods, such as LPs. Finally, a Monte Carlo analysis confirms that our empirical framework has no tendency to spuriously detect long-run effects when those are not present in the data-generating process.

Related Literature.—A large empirical literature has studied the macroeconomic effects of government spending on output over the business cycle. A key challenge is to isolate movements in public expenditure that are exogenous to economic conditions. Leading approaches have used narrative evidence (Ramey and Shapiro 1998), timing restrictions (Blanchard and Perotti 2002), sign restrictions (Mountford and Uhlig 2009), and geographical variation (Nakamura and Steinsson 2014; Chodorow-Reich 2019). In two comprehensive reviews, Ramey (2011a, 2019) summarizes the literature and concludes that the government spending multiplier lies between 0.6 and 1.5, across the reviewed papers. The focus on frequencies beyond the business cycle is a distinctive feature of our analysis.

An important strand of research has focused on the impact of public spending on productivity. Moretti, Steinwender, and Van Reenen (2025) and Deleidi and Mazzucato (2021) find that military expenditure fosters private innovation, while Gruber and Johnson (2019); Gross and Sampat (2023); Diebolt and Pellier (2020); and Ilzetzki (2024) document the long-lasting effects of the two world wars on US patenting and productivity. Kantor and Whalley (2023) show that the Space Race with the Soviet Union of the 1960s had persistent effects on manufacturing growth across US counties. Our historical analysis extends these event studies to a much longer sample and forecast horizon, using a different identification; furthermore, it shows that public R&D can stimulate productivity and output even in peacetime. This latter finding has been recently echoed by De Lipsis et al. (2022) and Fieldhouse and Mertens (2023), who report significant effects of post-WWII public R&D on US output and TFP.[2]

Our results also speak to the public infrastructure research surveyed by Ramey (2020). Fernald (1999) and Leff Yaffe (2020) find that the US interstate highway program boosted industry-level productivity, while Donaldson and Hornbeck (2016) and Hornbeck and Rotemberg (2021) estimate that the US national railroad network improved market access. We complement these studies by showing that public investment in equipment and structures tends to have smaller effects than public R&D at long horizons.

A growing literature, surveyed by Cerra, Fatás, and Saxena (2022), studies the long-lasting effects of demand shocks. Comin and Gertler (2006) and Beaudry,

[2] See Janeway (2012) and Mazzucato (2013) for earlier popular writings on the role of public spending in innovation.

Galizia, and Portier (2020) lay out models with strong internal propagation mechanisms in which nontechnology shocks have effects beyond the business cycle. Benigno and Fornaro (2018) focus on stagnation traps triggered by weak aggregate demand. Jordà, Singh, and Taylor (2020) exploit the international finance trilemma to identify the long-run effects of monetary policy. Akcigit et al. (2022) study the impact of income taxes on innovation and researchers' mobility across US states. Cloyne et al. (2025) estimate persistent responses of R&D, productivity, and GDP to corporate and personal tax changes. Our analysis offers a novel evaluation of the effects of government spending on the aggregate economy at long horizons.[3]

Structure of the Paper.—In Section I, we present our empirical framework, the historical data and the identification strategy. The main findings on productivity, output and the fiscal multiplier are reported in Section II, while, in Section III, we assess the robustness of our low-frequency inference using different samples, model specifications, and econometric methods. Exploring the transmission mechanism of fiscal policy working through the different categories of private and public spending is the focus of Section IV, whereas, in Section V, we contrast the large and long-lasting effects of public R&D shocks with the small and shorter-lived impact of public consumption and public investment innovations. Conclusions are discussed in Section VI. In the Supplemental Appendices, we provide details on the estimation and present further analyses.

I. Empirical Framework

In this section, we motivate the empirical model and the estimation strategy that we propose, including prior and lag length selection. We then present the historical data for the United States and review the identification of government spending shocks based on the military spending news constructed by Ramey (2011b) (which in turn builds upon Ramey and Shapiro 1998) and extended back in time by Ramey and Zubairy (2018). We complement their dataset with extended series for business investment, productivity, patents, consumption, exports, imports, and government spending broken down into its three main categories, including public R&D.

A. *Model Specification and Estimation*

We use a Vector Autoregressive (VAR) model to conduct inference on the effects of government spending on economic activities. The model can be written as

$$\mathbf{y}_t'\mathbf{A}_0 = \sum_{\ell=1}^{p} \mathbf{y}_{t-\ell}'\mathbf{A}_\ell + \mathbf{c} + \varepsilon_t' \quad \text{for } 1 \leq t \leq T, \tag{1}$$

[3] Following the Great Recessions of 2007–2009, an independent literature has shown that financial shocks can have long-lasting effects on the economy in business cycle models with financial frictions and endogenous total factor productivity. Prominent examples include Anzoategui et al. (2019); Bianchi, Kung, and Morales (2019); Guerron-Quintana and Jinnay (2019); Ikeda and Kurozumi (2019); and Queralto (2020), among many others.

where $\mathbf{y}_t$ is an $n \times 1$ vector of variables, $\boldsymbol{\varepsilon}_t$ is an $n \times 1$ vector of structural shocks, and $\mathbf{A}_\ell$ is an $n \times n$ matrix of parameters for $0 \leq \ell \leq p$, with $\mathbf{A}_0$ invertible. The vector of parameters $\mathbf{c}$ has dimension $1 \times n$ and the letter p refers to the lag length, whereas T denotes the sample size. The vector ε_t, conditional on past information and the initial conditions $\mathbf{y}_0, \ldots, \mathbf{y}_{1-p}$, is Gaussian with zero mean and covariance matrix $\mathbf{I}_n$, the $n \times n$ identity matrix.

Denoting $\mathbf{A}'_+ \equiv \left[\mathbf{A}'_1 \cdots \mathbf{A}'_p \mathbf{c}'\right]$, the reduced-form representation implied by equation (1) is $\mathbf{y}'_t = \sum_{\ell=1}^{p} \mathbf{y}'_{t-\ell}\mathbf{B}_\ell + \mathbf{d} + \mathbf{u}'_t$ for $1 \leq t \leq T$, or more compactly, $\mathbf{y}'_t = \mathbf{x}'_t\mathbf{B} + \mathbf{u}'_t$, where $\mathbf{x}'_t = \left[\mathbf{y}'_{t-1}, \ldots, \mathbf{y}'_{t-p}, 1\right]$, $\mathbf{B} = \mathbf{A}_+\mathbf{A}_0^{-1}$, $\mathbf{d} = \mathbf{c}\mathbf{A}_0^{-1}$, $\mathbf{u}'_t = \boldsymbol{\varepsilon}'_t\mathbf{A}_0^{-1}$, and $E\left[\mathbf{u}_t\mathbf{u}'_t\right] = \boldsymbol{\Sigma} = \left(\mathbf{A}_0\mathbf{A}'_0\right)^{-1}$. The matrices $\mathbf{B}$ and $\boldsymbol{\Sigma}$ are the reduced-form parameters, while $\mathbf{A}_0$ and $\mathbf{A}_+$ are the structural parameters. Similarly, $\mathbf{u}'_t$ are the reduced-form innovations, while $\boldsymbol{\varepsilon}'_t$ are the structural shocks. The shocks are orthogonal and have an economic interpretation, while the innovations are typically correlated and have no interpretation.

For computational simplicity and in order to preserve degrees of freedom, we assume time-invariant coefficients and Gaussian, homoskedastic innovations. Our macroeconomic data span 125 years and therefore are very likely to exhibit some form of time variation in parameters and volatilities (see, for instance, Sargent and Surico 2011). Given our interest on effects over horizons of up to 15 years, however, this leads to just 8 nonoverlapping samples, and hence, we refrain from any attempt to model time variation. Our impulse response estimates can thus be interpreted as averaging across different macroeconomic regimes over the sample. Furthermore, our impulse-response functions will be consistent even in the presence of heteroskedastic and non-Gaussian errors (Montiel Olea, Plagborg-Møller, and Qian 2022).

In the VAR setting, impulse-response functions (IRFs)—and related objects of interest, such as government spending multipliers, forecast error variance decompositions, etc.—are computed by recursively iterating on the VAR coefficients, $\boldsymbol{\Theta} = \left(\mathbf{A}_0, \mathbf{A}_+\right)$.[4] However, in recent years it has become increasingly popular to compute IRFs using direct regressions of the variable of interest in period $t + h$ on a measure of an identified shock at time t, as well as on control variables. As shown by Jordà (2005), these "local projections" can be written as

$$y_{i,t+h} = \alpha_h + \beta_h \hat{\varepsilon}_t^1 + \psi_h(L)\,\mathbf{z}'_t + \nu_{t+h} \quad \text{for } h = 0, 1, \ldots, H, \tag{2}$$

where $\hat{\varepsilon}_t^1$ is a proxy for the identified shock. For comparability and without loss of generality, we assume that the shock in the local projection (2) corresponds to the first shock in the VAR (1).

There has been considerable debate in the literature about the relative advantages of VAR versus local projection (LP) estimates of impulse responses.

[4] For instance, given a value $\boldsymbol{\Theta}$ of the structural parameters, the IRF of the i-th variable to the j-th structural shock at horizon k corresponds to the element in row i and column j of the matrix $\mathbf{L}_k(\boldsymbol{\Theta})$, defined recursively by

$$\mathbf{L}_0(\boldsymbol{\Theta}) = \left(\mathbf{A}_0^{-1}\right)', \quad \mathbf{L}_k(\boldsymbol{\Theta}) = \sum_{\ell=1}^{k}\left(\mathbf{A}_\ell\mathbf{A}_0^{-1}\right)'\mathbf{L}_{k-\ell}(\boldsymbol{\Theta}), \text{ for } 1 \leq k \leq p,$$

$$\mathbf{L}_k(\boldsymbol{\Theta}) = \sum_{\ell=1}^{p}\left(\mathbf{A}_\ell\mathbf{A}_0^{-1}\right)'\mathbf{L}_{k-\ell}(\boldsymbol{\Theta}), \text{ for } p < k < \infty.$$

Plagborg-Møller and Wolf (2021); Montiel Olea and Plagborg-Møller (2021); and Li et al. (2021b) clarify important conceptual and practical aspects and conclude that the two approaches estimate the same impulse responses in population. In particular, their estimands approximately coincide up to horizon p (the maximum lag length of the VAR). Furthermore, standard confidence intervals based on lag-augmented LP have correct asymptotic coverage, uniformly, over the persistence in the data-generating process and over a wide range of horizons. Finally, in small-sample applications, a trade-off emerges between the higher bias of low-order VARs and the higher variance of LPs, such that shrinkage estimators—e.g., Bayesian VARs or penalized LPs (Barnichon and Brownlees 2019)—become attractive.[5] In our context, with nonstationary variables and cointegrating relationships, Bayesian VARs are an effective tool to address the finite sample bias that characterizes autoregressions containing unit roots via priors elicited on the system as a whole (Doan, Litterman, and Sims 1984; Sims, Stock, and Watson 1990; Sims 1993; Sims and Zha 1998; Giannone, Lenza, and Primiceri 2015, 2019). This compares favorably with single-equation methods like LPs.

Our focus on low frequencies requires a careful consideration of the small sample bias-variance trade-off highlighted by Li, Plagborg-Møller, and Wolf (2021a). To balance these two considerations, we set the lag length of our baseline VAR to $p = 60$. The rationale for this choice is twofold. First, we want to look at horizons well beyond the eight years traditionally associated with business cycle frequencies. Second, we are interested in capturing potentially long lags in the diffusion of technological advances after a surge in R&D spending.

As for inference, we take a Bayesian approach and apply priors that shrink coefficients toward zero at a rate that exponentially increases with the more distant lags, in the spirit of the "Minnesota" priors of Doan, Litterman, and Sims (1984) and Sims (1993). The generous choice of lag length brings the impulse responses of the VAR close to what would have been obtained with lag-augmented LPs, whereas the use of shrinkage allows us to mitigate the increase in variance stemming from the very large number of parameters involved. It is worth noting that the Minnesota priors place a heavier shrinkage on more distant lags (centered around the value of zero), and therefore, the data need to speak strongly about the presence of low-frequency dynamics to counteract the a priori view that these are likely absent. Further details on the specification of the prior are given below.

B. *Prior Specification and Posterior Sampling*

We will use a Normal-Inverse Wishart prior over the reduced-form parameters, $(\mathbf{B}, \boldsymbol{\Sigma})$. This family of distributions is conjugate for this class of models and is the standard choice in empirical work due to its computational tractability (see, for instance, Uhlig 2005; Giannone, Lenza, and Primiceri 2015). Denoting $\mathbf{b} = vec(\mathbf{B})$, the prior distribution is $NIW(\underline{\nu}, \underline{\boldsymbol{\Psi}}, \underline{\mathbf{b}}, \underline{\mathbf{V}})$ As discussed above, we employ the "Minnesota" priors proposed by Doan, Litterman, and Sims (1984), which shrink the VAR coefficients toward simple univariate specifications. In particular, the degrees

[5] Penalized LPs minimize the sum of squared forecast errors plus a penalty term that encourages IRF smoothness.

of freedom of the prior covariance matrix are set to $\underline{\nu} = n + 2$, with $\underline{\Psi}$ a diagonal matrix whose j-th diagonal element is ψ_j.[6] As for the autoregressive coefficients, the prior has the following mean and variance:

$$E\left[(\mathbf{B}_\ell)_{i,j}|\boldsymbol{\Sigma}\right] = \begin{cases} \delta, & \text{if } j = 1 \text{ and } \ell = 1 \\ 0, & \text{otherwise} \end{cases} \tag{3}$$

$$\operatorname{cov}\left[(\mathbf{B}_\ell)_{i,j}, (\mathbf{B}_m)_{r,k}|\boldsymbol{\Sigma}\right] = \begin{cases} \lambda^2 \frac{1}{\ell^\alpha} \frac{\Sigma_{i,r}}{\psi_j/(\underline{\nu} - n - 1)}, & \text{if } j = k \text{ and } \ell = m \\ 0, & \text{otherwise} \end{cases}. \tag{4}$$

The parameter δ, which is the mean of the autoregressive coefficient corresponding to the first lag, is set to 1 for trending variables, to 0.9 for stationary but persistent variables, and to 0 for other variables. As discussed by Del Negro and Schorfheide (2011), among others, the hyperparameter λ controls the overall tightness of the Minnesota prior, whereas the term $1/\ell^\alpha$ implies that more distant lags are shrunk at an exponentially increasing rate toward zero, with the hyperparameter α determining how aggressively longer lags are penalized. Therefore, the Minnesota prior penalizes rich large structures and favors models with shorter lags and "smooth" impulse responses.

Because our dataset contains a mix of stationary and nonstationary variables, we combine the Minnesota prior with the "Single Unit Root" prior proposed by Sims (1993) and Sims and Zha (1998). This prior addresses the problem of the excessive explanatory power of initial conditions and deterministic components, which translates into downward bias in the persistence of autoregressive coefficients (see Sims and Uhlig 1991; Sims 2000; Jarociński and Marcet 2014; Giannone, Lenza, and Primiceri 2019). It is usually implemented by appending an artificial ("dummy") observation for $\mathbf{y}$ and $\mathbf{x}$, denoted y_* and x_*, respectively, at the beginning of the sample:

$$\mathbf{y}^*_{1\times n} \equiv \frac{\bar{y}}{\theta}; \quad \mathbf{x}^*_{1\times(n\cdot p+1)} \equiv \left[\frac{1}{\theta}, y_*, \ldots, y_*\right], \tag{5}$$

where $\bar{y}$ is the average of the first p observations and the hyperparameter θ controls the tightness of the prior. A smaller θ implies a tighter prior in favor of unit roots and cointegration in the system as a whole, inducing a priori correlation between the constant and the different lags of the VAR.[7] This combination of priors is widely used in empirical macroeconomics. The conjugate nature of the prior allows us to sample from the posterior distribution in a straightforward way, using the standard algorithm described in the Supplemental Appendix.

In our context where the number of parameters is large relative to the sample size, the choice of prior hyperparameters might become important for the posterior impulse responses. In particular, if λ or θ are large (or α is small), the priors are too

[6] As common, we set $\underline{\Psi}_{j,j} \equiv \psi_j$ to the residual variance of a univariate AR (1) estimated on the full sample.

[7] The likely presence of cointegration in our dataset leads us to not use in our baseline results the other well-known prior used in the empirical macroeconomics literature, known as the "sum of coefficients" prior. See the discussion in Giannone, Lenza, and Primiceri (2019). We explore the role of this prior in Section IIIB.

loose, and the large number of parameters means that IRFs will be estimated imprecisely. On the other hand, as $\lambda, \theta \rightarrow 0$ and/or $\alpha \rightarrow \infty$, medium-term dynamics may be smoothed away by the priors, similar to using a smaller amount of lags. Giannone, Lenza, and Primiceri (2015) propose a theoretically grounded method to optimally choose the prior hyperparameters, based on maximization of the marginal likelihood. Based on this procedure, we select $\lambda = 0.36$, $\alpha = 2$, and $\theta = 0.01$ for our baseline estimates. In Section IIIB and the Supplemental Appendix, we explore in detail the impact of the priors specification on the empirical results, whereas in the Supplemental Appendix we assess the sensitivity of the marginal likelihood to different hyperparameter choices.

C. *Data and Identification*

Our starting point is the dataset in Ramey and Zubairy (2018), which spans the sample 1890:I to 2015:IV and contains the present discounted value of military news (Ramey 2011b), government spending, real GDP, the log GDP deflator, the short-term interest rate, the surplus-to-GDP ratio, and the Debt-to-GDP ratio. In drawing inference about low frequencies, our baseline approach is to express nonstationary variables in log-levels. Sims, Stock, and Watson (1990) show that, even in the presence of cointegration, this specification leads to consistent estimates. When computing government spending multipliers, however, the log-level specification requires scaling the impulse responses by the steady-state value of Y/G. As discussed by Owyang, Ramey, and Zubairy (2013) and Ramey and Zubairy (2018), multiplier estimates can be quite sensitive to this conversion factor measured from historical averages. Accordingly, we also compute output multipliers from alternative models in which GDP and government spending are scaled either by GDP in the previous quarter (as in Barro and Redlick 2011) or by the measure of GDP trend proposed by Ramey and Zubairy (2018): a sixth-degree polynomial for log GDP, 1889:I–2015:IV, excluding 1930:I–1946:IV. The baseline transformation includes an intercept and thus implicitly controls for a linear trend; the second transformation is akin to estimating the VAR in differences, hence removing a stochastic trend; the third transformation has the disadvantage of purging low-frequency movements in potential output that may be particularly important to account for highly persistent effects of government spending. We include it mostly for comparability with the estimates in Ramey and Zubairy (2018).

We extend the baseline data along several dimensions. First, we construct new quarterly series of private consumption and investment, as well as exports and imports. We obtain unpublished annual estimates of investment since 1901 by the Bureau of Economic Analysis. Before that, we rely on the Macrohistory Database of Jordà, Schularick, and Taylor (2017), which also offers a measure of annual private consumption since 1890. We interpolate these series to quarterly frequency using the consumption and investment series from NIPA (after 1947), Gordon (2007) (between 1919 and 1940), and real GDP (before 1919 and from 1941 to 1946), and use import and export data to ensure consistency of our estimates with the national accounts identity. Second, we construct quarterly measures of hours worked and TFP. The annual hours and productivity series comes from Bergeaud, Cette, and Lecat (2016). We adjust TFP for capital and labor utilization following Imbs (1999).

We interpolate this using the quarterly series of adjusted TFP in Fernald (2012) (after 1947) and real GDP (before 1947). The data on patents are by IFI CLAIMS Patent Services via Google Public Data.

In addition, we construct new historical series of public consumption and investment, distinguishing between expenditure in Equipment and Structures (E&S) and in Research and Development (R&D). Official NIPA estimates start in 1929. We reconstruct the series of public investment and its components for the period 1890–1929 by digitizing detailed government outlays data from both the *Historical Statistics of the United States* (Bureau of the Census 1949) and the annual *Statistical Abstracts* published by the census. We rely on the narrative evidence in Bush (1954) and Dupree (1986) to classify investment into E&S and R&D. Finally, we interpolate the resulting annual series using quarterly government spending and back out public consumption as residual. Further details on the construction of all series are provided in the Supplemental Appendix.

When moving from annual to quarterly frequency, we use the method by Chow and Lin (1971). It is worth emphasizing that the impulse responses at long horizons, which are the primary focus of our analysis, depend mainly on the low-frequency properties of the data, which in turn are pinned down by the properties of the annual series. With the exception of the reconstructed government spending series, these annual series are mostly available from existing sources, which we take at face value. The interpolation affects mostly the high-frequency properties of the data (i.e., within the year), and, as such, it seems unlikely to have an effect on the estimated IRFs at longer horizons. We verify this hypothesis in Section IIIA.

To identify the structural parameters of the VAR, we follow the approach labeled as "internal instruments" by Plagborg-Møller and Wolf (2021) and also used by Ramey (2011b). This approach includes the instrumental variable (in our case the military spending news series) in the VAR and identifies the shock of interest by ordering the instrument first in a Cholesky decomposition. This approach is attractive because it will automatically control for any residual predictability contained in the instrument and still yield valid impulse responses when the instrument is contaminated with measurement error that is unrelated to the shock of interest.[8]

II. The Effects of Military Spending

In this section, we report our main results, which are based on a quarterly VAR with 60 lags and the following variables (described in the previous section): military news, government spending, real GDP, adjusted total factor productivity, the short-term interest rate, the surplus-to-GDP ratio, and the debt-to-GDP ratio.[9] We begin by analyzing the impulse responses to a military spending shock and then move to the estimates of the output multipliers across forecast horizons, up to 60

[8] Plagborg-Møller and Wolf (2021) point out that this approach yields valid impulse response estimates even if the shock of interest is noninvertible. However, in presenting estimators such as the Forecast Error Variance Decomposition (see Supplemental Appendix), and in the alternative identification based on the maximum share of the variance (Section VB), we will require invertibility and no measurement error.

[9] Relative to the seven-variable VAR in Ramey and Zubairy (2018), we have replaced GDP deflator with TFP, as the latter is central to the transmission mechanism highlighted in this paper. But, in Antolin-Diaz and Surico (2022), we have verified that our VAR (60) estimates very similar effects at long horizons using their original set of 7 variables.

quarters. In the next section, we assess the reliability of our low-frequency inference by presenting an extensive set of robustness checks, evaluating the role of the priors, conducting Monte Carlo analyses, and reporting frequentist estimates of local projections. In Section IV, we present the results of an extended VAR, where we add newly constructed time series of consumption, investment, trade, patents, and the three main components of government spending since 1890:I to shed light on the transmission of fiscal shocks.

A. *Impulse Response Analysis*

A simple way to summarize the estimates of a VAR is to report impulse responses of the endogenous variables to the identified shock of interest. We select a forecast horizon of 60 quarters to match the number of lags chosen in the estimated VAR (60) and report pointwise 68 percent and 90 percent posterior credible sets (as shaded areas). For ease of interpretation, the military spending news shock is normalized so as to increase government spending by 1 percent of GDP over the first year after the shock. The top row of Figure 1 presents the responses of government spending and real GDP. The middle row refers to the short-term nominal interest rate and TFP, whereas the bottom row focuses on the government balance sheet: fiscal deficit and public debt, both expressed as a share of GDP.

The main findings from our VAR (60) can be summarized as follows. During the first four years after the shock, government spending increases sharply and then reverts, triggering an equally persistent increase in GDP, a notable fiscal deterioration with government debt peaking around 1.5 percent of GDP, and a delayed but significant increase in productivity. At frequencies between five and eight years, government spending goes back to its initial level, causing a short-lived slowdown in both output and TFP. This is associated with a switch toward fiscal surplus that contributes to revert the path of the debt-to-GDP ratio.[10]

In the long term, conventionally defined as frequencies beyond eight years, the response of government spending becomes significant again, but its peak is now a fraction of what was at shorter horizons. The fiscal surplus is no longer statistically different from zero, and public debt slowly returns to pre-shock levels. In contrast, GDP and total factor productivity witness a second boom that is not only as large in magnitude as the first peak but also appears more persistent. Interestingly, the timing of the productivity response is consistent with the empirical literature on the rate of technological diffusion, which typically estimates adoption lags between 6 and 17 years (Comin and Mestieri 2014; Pezzoni, Veugelers, and Visentin 2022).

There is some tentative evidence that the effects on output and TFP might weaken somehow after 15 years. It should be noted, however, that given the large number of lags and the long forecast horizon (relative to the sample size), caution should be exercised in claiming that empirical analyses such as ours could possibly distinguish

[10]The sequence of fiscal surpluses from year 4 to 10 in Figure 1 are notably smaller than the fiscal deficits triggered by the initial government spending expansion. This suggests that the (second wave of) GDP response plays a major role in reducing the debt-to-GDP ratio to pre-shock levels, consistent with the evidence in Hall and Sargent (2011).

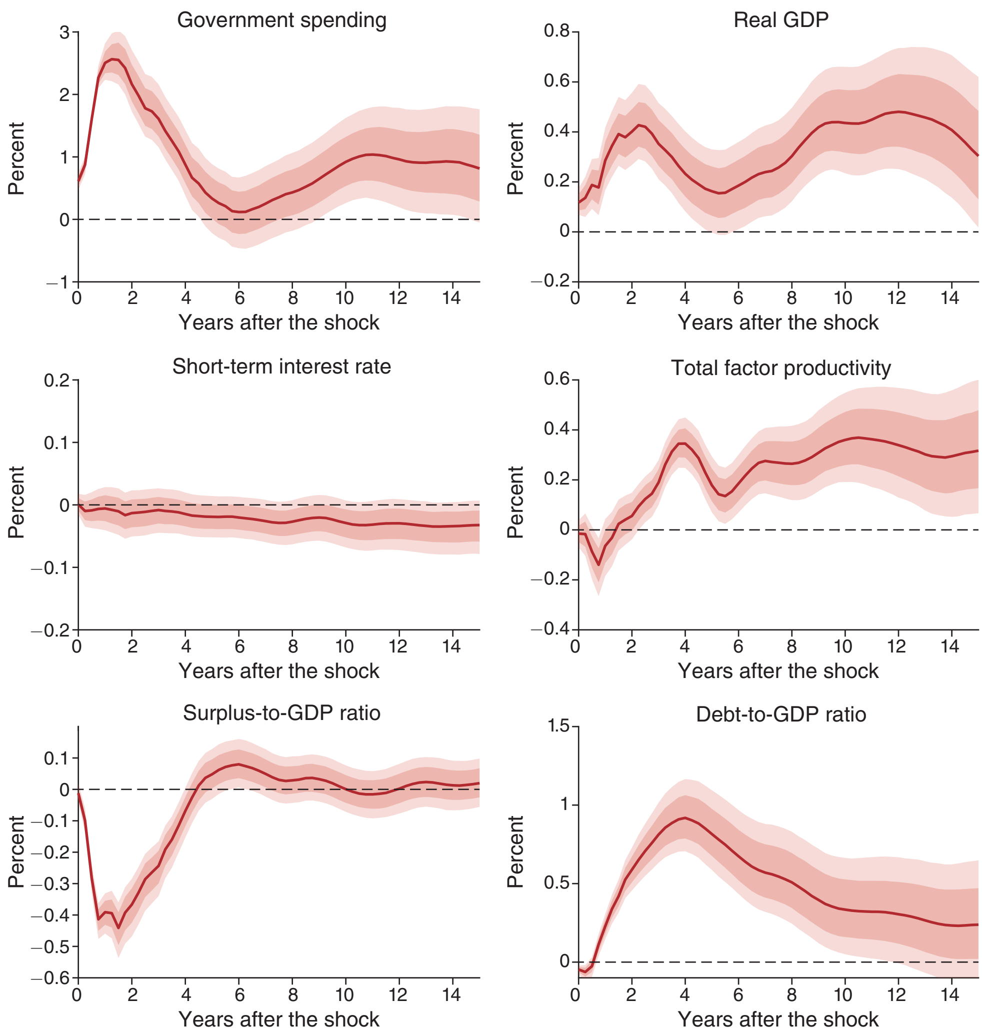


FIGURE 1. IMPULSE RESPONSES TO MILITARY NEWS SHOCK

Notes: The impulse responses are based on an estimated VAR with 60 lags of military spending news, government spending, real per capita GDP, utilization-adjusted TFP, short-term interest rate, fiscal deficit to GDP ratio, and government debt to GDP ratio. Government spending, GDP, and TFP enter the VAR in log-levels. Military spending news is ordered first in the Cholesky factorization. The size of the shock is normalized such as to increase government spending by 1 percent of GDP over the first year after the shock, based on a median G/Y ratio of 19. The darker (lighter) shaded areas represent the central 68 percent (90 percent) high–posterior density (HPD) intervals. The darker solid lines are the median estimates. Results are based on 5,000 posterior draws.

between truly permanent and very persistent dynamics.[11] Our preferred interpretation of the evidence in Figure 1 (and the rest of our analysis) is that the effects of government spending on output and TFP are very likely to extend beyond business

[11] Forecasting 15+ years ahead using 125 years of data relies on fewer than 8 nonoverlapping samples of 15 years.

cycle frequencies. Finally, the effects on the short-term nominal interest rate are negligible throughout.[12]

For completeness, we report the Forecast Error Variance Decomposition (FEVD) in the Supplemental Appendix. Military spending shocks explain between 30 percent and 40 percent of the variation in government spending at business cycle frequencies, and about 20 percent of fluctuations at longer horizons. These shocks account for a nontrivial fraction of the variance of GDP and productivity, around 10 percent. This is consistent with the evidence in Rossi and Zubairy (2011) on the role of fiscal policy in explaining US medium-term fluctuations.

In summary, we estimate significant long-lasting effects of government spending on both output and productivity. Unlike the short-run dynamics where the movements in government spending tend to be of a similar magnitude (if not larger) than the response of GDP, the lower frequency estimates suggest a large multiplier at long horizons, as the effects on output are associated with far smaller changes in government spending at longer horizons. In the next part of this section, we corroborate this conjecture by formally computing the multiplier across forecast horizons.

B. *The Government Spending Multiplier in the Short Run and the Long Run*

In the previous section, we have estimated a larger output response at longer horizons relative to the smaller lower-frequency movements in government spending, and the opposite at higher frequencies. In this section, we formally quantify these relative effects by computing the fiscal multiplier of government spending on output across forecast horizons. This is interesting for at least two reasons. First, government spending may have different effects at different horizons, and comparing the multipliers at high, business cycle, and low frequencies within the same estimated model can help shed light on this issue. Second, as noted by Ramey (2019), different studies often compute the multiplier at different horizons, and reporting how the estimates of this statistics vary with the forecast horizon may help reconcile seemingly conflicting findings in the literature.

In line with earlier work, we define the output multiplier for each horizon h as the ratio between the cumulative impulse response of real GDP to military spending news up to horizon h and the cumulative impulse response of government spending to the same shock over the same horizon. Following Mountford and Uhlig (2009), we use the sample average nominal interest rate to discount the estimates between one and h quarters ahead. In Figure 2, we display the present value multiplier for each horizon between $h = 0$ (i.e., the impact multiplier) and $h = 60$ (i.e., the long-run multiplier). Panel A refers to the specification in log-levels and uses the historical median of $G/Y = 19\%$ to transform the estimated elasticities into multipliers. Panel B refers to the specification in which output and government spending are both scaled by Y_{t-1}. Panel C is based on a model where both government

[12] Using the yield on 10-year government bonds instead of the short-term rate in the VAR produces very similar findings. As noted by Meltzer (2004), until the Treasury-Fed accord of 1951, the Fed pegged interest rates at a low level to facilitate the financing of government debt during wartime. Friedman and Schwartz (1963) argue that the Fed choice of not controlling the growth of the monetary base over this period contributed to fueling inflation. This is consistent with the responses of prices in Antolin-Diaz and Surico (2022) and interest rates in Figure 1, respectively.

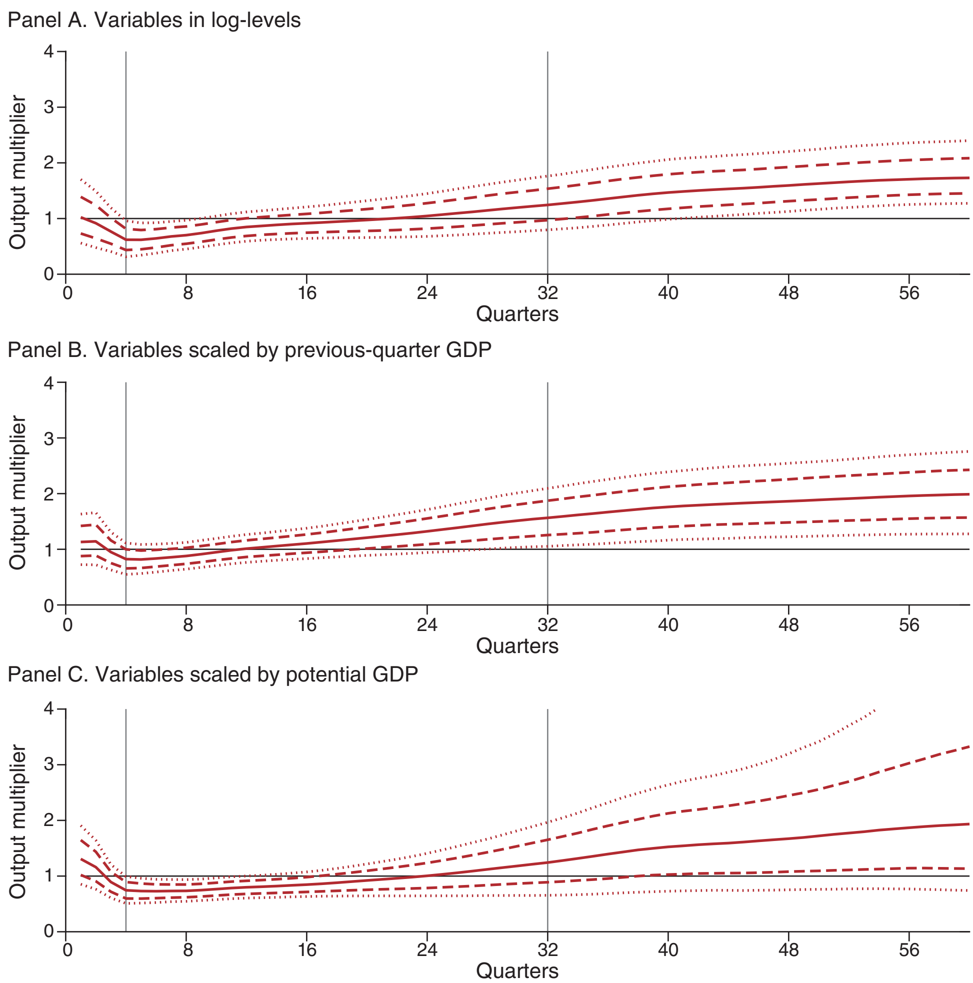


Figure 2. The Government Spending Multiplier across Horizons

Notes: The present value multiplier at each horizon h is computed as the ratio of the integral up to horizon h of the output response and the integral up to horizon h of government spending response to a military spending news shock, discounted using the steady-state interest rate. The estimates are based on VARs with 60 lags. In the top panel, government spending and output enter the VAR in log-levels, and the multipliers are obtained using the elasticity formula and the historical median G/Y ratio of 19 percent. In the middle panel, output and government spending are both divided by Y_{t-1}. In the bottom panel, they are expressed in percent of potential output as defined in Ramey and Zubairy (2018). The broken (dotted) lines represent the central 68 percent (90 percent) HPD interval. The solid line stands for the median estimate. Results are based on 5,000 posterior draws.

spending and real GDP are scaled by potential output, as defined by Ramey and Zubairy (2018). The latter two strategies provide direct estimates of the multipliers and do not rely on the government spending-output ratio.

The estimates in Figure 2 reveal that the posterior distributions of the government spending multiplier display, on impact, median values between 1 and 1.35, with most mass above 1. After the first year, however, the multiplier decreases below 1, consistent with the evidence in Hall (2009); Barro and Redlick (2011); and Ramey and Zubairy (2018). These estimates are relatively stable over the following three to five years before growing with the forecast horizon. The posterior median takes values above 1 at frequencies beyond 8 years, and it peaks at values between 1.7 and 2 (across specifications) in the forecasts 15 years ahead. Interestingly, despite relatively close median estimates, both the log-level and the previous-quarter-GDP specifications lead to more accurate inference about the long-run multiplier than the model that removes potential output.

In summary, the findings of this section suggest two main conclusions about the effects of government spending on output. First, on impact and at business cycle frequencies (i.e., from 6 to 32 quarters) the multipliers span the range of point estimates available in the fiscal policy literature, between 0.6 and 1.5, thereby offering a possible reconciliation of apparently conflicting results in earlier empirical macro studies. Second, while the multipliers at business cycle frequencies tend to exhibit values below or around 1, the multipliers at low frequencies (i.e., beyond 32 quarters) display much larger values and eventually exceed 1 significantly at long horizons.

III. Assessing Inference at Low Frequencies

In the previous section, we have reported strong evidence of significant effects of government spending on output and productivity at horizons beyond business cycle frequencies (i.e., after eight years). In this section, we want to assess the robustness of our findings to several modifications of our sample, model specification, and econometric method. We start by reporting results for samples that exclude from the instrument either WWII or one cluster of military spending events at the time to make sure no single episode drives the identification. Then, we look at IRFs based on VARs estimated using annual data (the frequency of some of our primary sources) or a considerably shorter lag length. Next, we move to assess the role of the priors by (i) modifying either their mean or variance, (ii) estimating jointly the prior hyperparameters using hierarchical priors, and (iii) applying the methods developed by Müller (2012) to measure both the sensitivity and informativeness of the priors. Then, we conduct a Monte Carlo analysis under two very different data-generating processes featuring, respectively, fully i.i.d. and nonstationary time series; the goal is to evaluate whether our empirical model has any tendency to spuriously detect long-lasting effects when these are actually *not* present in the data-generating process. Finally, we present frequentist estimates based on local projections, which have lower bias but higher variance at long horizons than the estimates based on VARs with short lag length (Li, Plagborg-Møller, and Wolf 2021b). All the analyses in this section corroborate, by and large, the notion that the effects of government spending extend significantly beyond business cycle frequencies.

A. *Sensitivity Analysis*

In this section, we check the robustness of our results to different samples, data frequency, and lag length selection. We record these results in Figure 3, whose columns refers to output and TFP, respectively. The top two rows refer to samples that exclude from the instrument either WWII (first row) or 12 major clusters of military spending news events one at a time (second row). As argued by Friedman (1952),

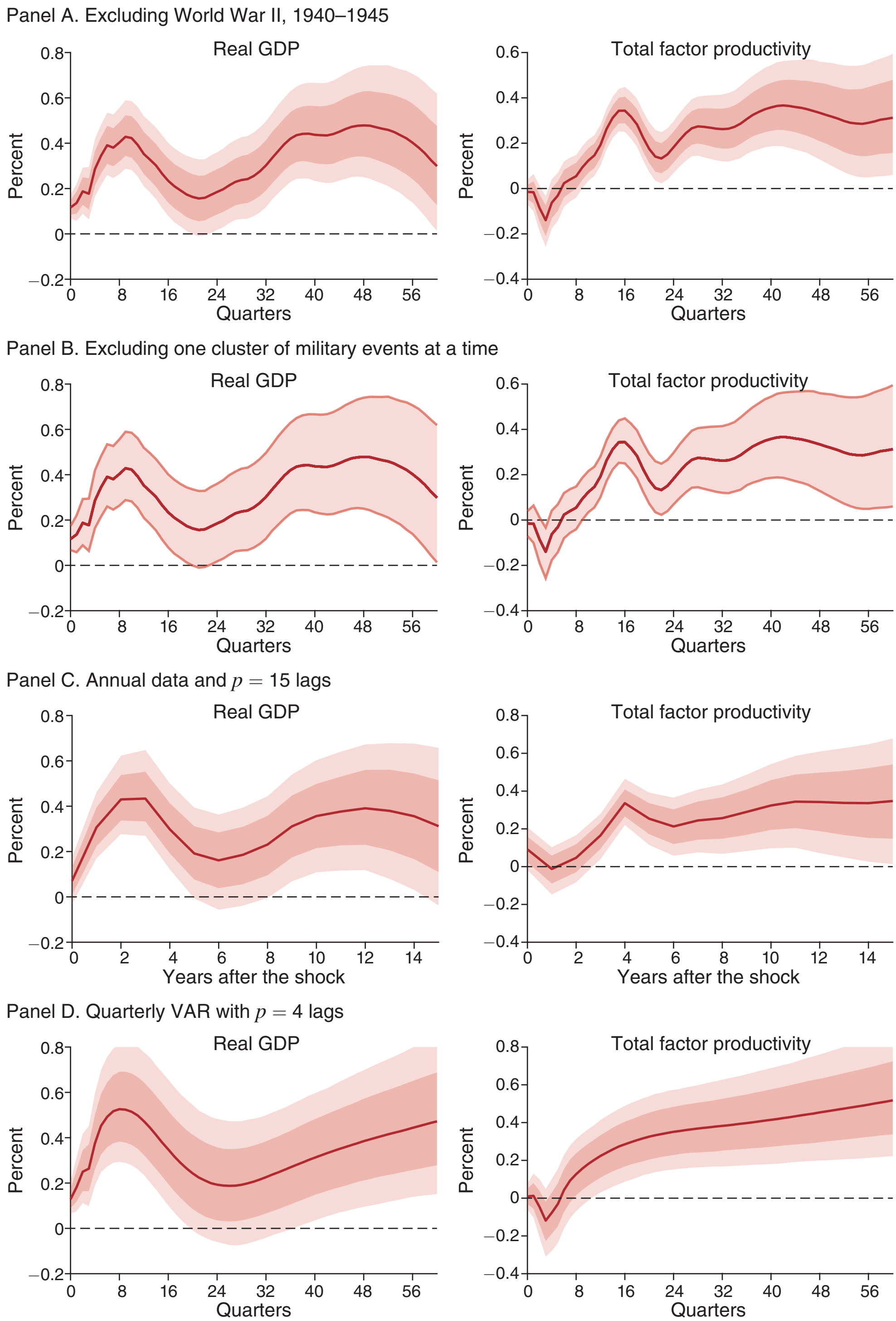


Figure 3. Sensitivity Analysis

Notes: The solid lines represent the median posterior response. The darker shadow area represents the sixty-eighth posterior credible intervals, while the lighter shadow area represents the ninetieth posterior credible intervals. Results are based on 5,000 posterior draws.

exploiting large wars and military spending for identifying the effects of government expenditure is attractive for at least two reasons. First, the variation in military spending associated with wars (abroad) is typically independent from the state of the (domestic) business cycle and thus should prevent reverse causality running from GDP to government spending. Second, public spending swings tend to be large in historical perspective, thereby offering sufficient variation in the leading variable. On the other hand, using wars as source of exogenous variation poses the external validity challenge that a specific episode may be driving the results. This concern is particularly acute for WWII, which represents—by far—the largest increase of real government spending in US economic history, both relative to GDP and in absolute value.

Excluding the years from 1940 to 1945, in the first row of Figure 3, produces IRFs for output and productivity that are very similar to the estimates in Figure 1. The response of GDP to a military spending shock occurring outside the WWII period is significant over the first 4 years, slows down between 16 and 32 quarters, and then increases again, for a longer period, at frequencies beyond year 8. The first peak of TFP is also very significant but appears delayed relative to the peak in GDP. The increase in productivity decelerates in years 5 and 6 before taking off again, persistently, at intermediate and long horizons. In the second row of Figure 3, we report the envelope of 90 percent credible intervals for 12 different exercises in which we have removed—one at a time—each major cluster of military spending.[13] The swath of median posterior estimates for GDP and TFP are very close to their Figure 1 counterparts. The envelope of 90 percent credible sets is somewhat larger than in the baseline results, but it is worth emphasizing that our main finding of large and significant effects of government spending at frequencies beyond the business cycle is not overturned by the removal of any of these events.[14] We also discuss subsample stability in the Supplemental Appendix.

In the third row of Figure 3, we come to terms with the fact that many of the variables we have reconstructed for the quarterly analysis have been interpolated, as historical data are typically reported at annual frequency by most primary sources. In Section IIA, we have discussed the reasons why this is unlikely to pose a threat to our low-frequency estimates. Here, we wish to verify that argument by running our model on annual data. For consistency with the quarterly analysis, we reduce the number of lags to 15 and adapt the priors to embed the same degree of persistence

[13] Historically, military news shocks tend to cluster around major wars and significant historical events. Accordingly, in the second row of Figure 3, we report estimated impulse responses based on subsamples in which we have removed one cluster of military spending at a time from the instrument. This is a more stringent test than simply removing one observation at a time. The clusters are June 1890 (Navy Bill), June 1898 to September 1898 (Cuban War), December 1915 to December 1918 (World War I), June 1940 to December 1945 (World War II), September 1950 to December 1953 (Korean War), December 1957 (Sputnik), March 1961 to December 1961 (Kennedy era), March 1965 to December 1967 (Vietnam War), March 1980 to March 1981 (Cold War buildup), December 1986 to March 1992 (end of Cold War), September 2001 to September 2008 (War on Terror), and December 2008 to December 2015 (Afghanistan War surge).

[14] We have also verified that the exclusion of any pair of events among the three largest war-induced military spending episodes (namely, WWI, WWII, and the Korean War) does not overturn our main conclusions: The long-run effects on output and productivity are still large and significant. On the other hand, excluding all these three large war episodes at once produces small and insignificant output and productivity responses at longer horizons. In other words, each and every one of these war-induced, large increases in government expenditure seems sufficient to elicit significantly persistent effects on the US economy in the long run, though none of them is actually necessary.

relative to the priors of the quarterly model. The estimates for output and TFP are very similar to the IRFs in Figure 1. The output response is characterized by two peaks, with a more persistent effect after year 8, and the productivity response is delayed, with a persistent increase at intermediate and long horizons.

In the fourth row of Figure 3, we ask whether a VAR with only four lags, a standard choice in most empirical macro analyses on quarterly data, is capable of fully capturing the dynamic responses of GDP and TFP. On the one hand, the evidence from the VAR (4) points to large and significant effects of military spending on output and productivity at frequencies beyond the business cycle. On the other hand, the estimated dynamics are—by construction—much smoother, and the effects look now even more persistent. As discussed in Section IIA, however, we stress that given the large number of lags, long forecast horizon and relatively short sample, the reader should resist the temptation to draw inference based upon whether the effects reported in this paper are best viewed as permanent or very persistent. Notwithstanding this interpretation caveat, all analyses in this paper point toward large and significant effects of government spending on GDP and TFP beyond business cycle frequencies. This is our favorite interpretation of our main findings.

B. *The Role of the Priors*

In this section, we analyze the impact of our Bayesian priors on our main empirical results. We are particularly interested in confirming that it is the information contained in the likelihood, rather than some feature of the priors, that drives our finding of significant effects of military spending on GDP and TFP beyond business cycle frequencies. An additional concern is that because the Minnesota priors shrink some series toward being random walks and others toward being persistent, stationary processes, and because the single unit root priors favor cointegration in the system, we might be building in, a priori, a bias toward finding very long-lasting effects when, in fact, these are transitory.

In Figure 4, we report a number of alternative specifications of the prior distributions, which are aimed at understanding the impact of their structure on the results. In panel A, we remove the single unit root prior, while in panel B, we additionally set the prior mean of the first autoregressive coefficient, δ, equal to zero for all variables. The latter choice can be regarded as a misspecification of the priors, as variables in level such as real GDP are known to be characterized by trending behavior. A further exercise is presented in panel C, where we introduce the "Sum of Coefficients" prior originally proposed by Doan, Litterman, and Sims (1984).[15] Again, the use of this type of prior might be regarded as a source of misspecification, as it introduces a bias *against* the presence of cointegration (see the discussion in Giannone, Lenza, and Primiceri 2019). The robust message delivered by the first three rows of Figure 4 is that the posterior results are qualitatively very similar to the estimates in Figure 1. We interpret this finding as evidence that no specific feature of

[15] Similar to the single unit root specification, this prior can be implemented by the addition of a series of dummy observations stacked on top of the data matrix, in particular $\mathbf{y}^{**}_{n\times n} \equiv \mathrm{diag}(\bar{y}/\mu); \mathbf{x}^{**}_{n\times(n\cdot p+1)} \equiv [0, y_{*}, \ldots, y_{*}]$, with μ being a hyperparameter controlling the tightness of the prior and $\bar{y}$ the average of the first p observations.

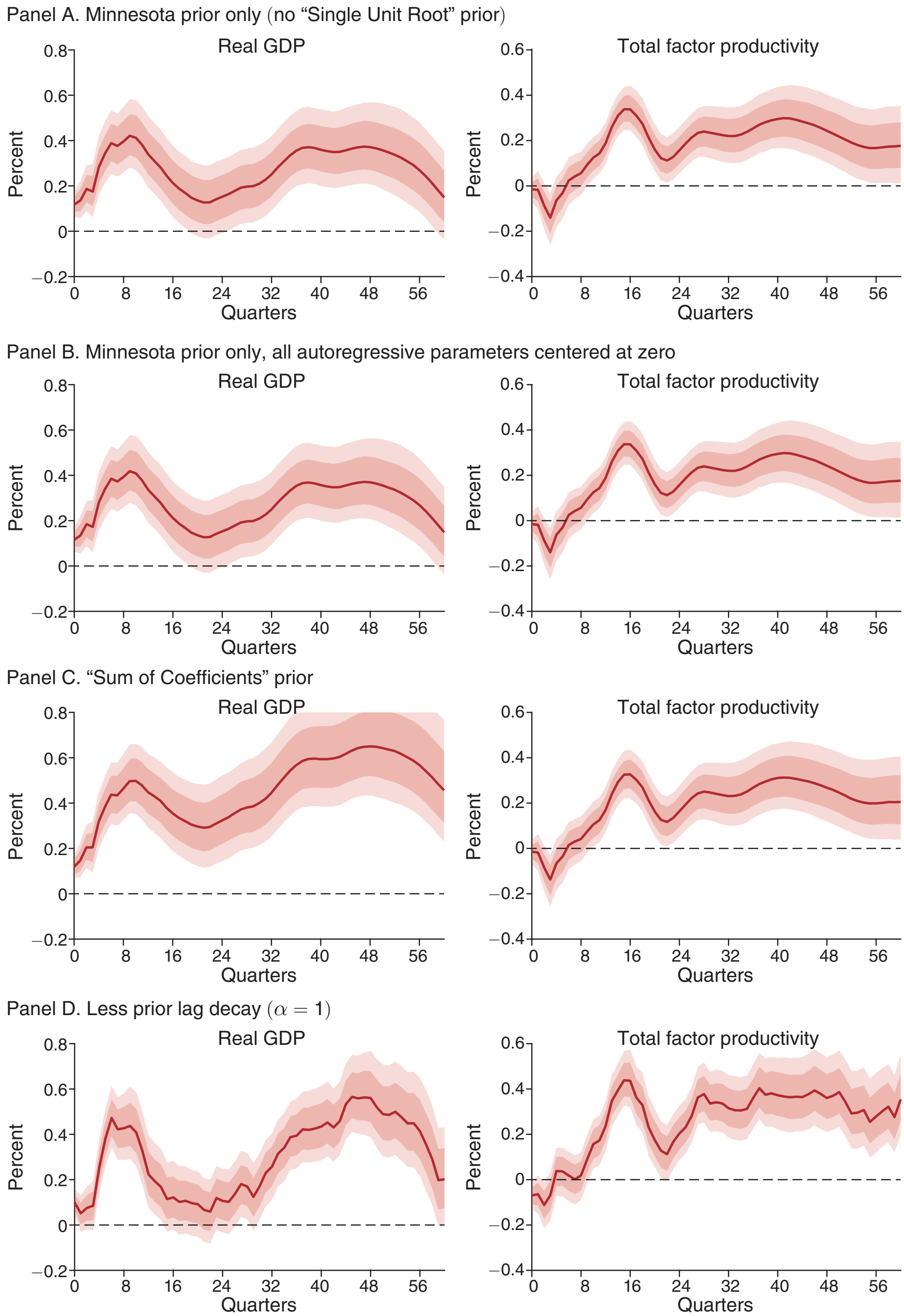


FIGURE 4. IMPULSE RESPONSE FUNCTIONS UNDER ALTERNATIVE PRIORS

Notes: The solid lines represent the median posterior response. The darker shadow area represents the sixty-eighth posterior credible intervals, while the lighter shadow area represents the ninetieth posterior credible intervals. Results are based on 5,000 posterior draws.

our baseline priors is responsible for the finding of significant effects of government spending on GDP and TFP at horizons beyond business cycle frequencies.

Another important question concerns whether the aggressive prior lag decay implied by the Minnesota prior could induce the empirical autocovariances at long lags ($\ell \approx 60$) to have only a minor influence on the posterior inference about impulse responses at long horizons. To address this issue, we also examine the alternative strategy of setting a linear, rather than exponential, decay pattern, that is, setting the hyperparameter $\alpha = 1$. As seen in panel D of Figure 4, the IRFs of GDP and TFP become substantially less smooth, but if anything, the results become stronger. Therefore, one can interpret our baseline specification of the priors as conservative, in the specific sense that our baseline choices tend to push the impulse responses at long horizons toward zero.

While the above exercises are reassuring, they still have the drawback that we are examining the isolated impact of changing features of the priors one at a time. We present below two exercises aimed at understanding the impact of the prior and its different hyperparameters altogether. First, following Giannone, Lenza, and Primiceri (2015), rather than searching for the value of the prior hyperparameters that maximizes the marginal likelihood, we simulate them from their full posterior distribution, allowing us to account for their estimation uncertainty.[16] This approach, which has a natural interpretation of a Bayesian hierarchical model, is implemented using a Metropolis step to draw the low-dimensional vector of hyperparameters. Using this alternative specification, in Figure 5, we present IRFs that are the counterparts of the estimates in Figure 1. Two main findings stem out from this exercise. First, the posterior credible sets of all impulse responses are wider than in the baseline case. This is not surprising because the hierarchical prior structure integrates across possible values of all hyperparameters, and this introduces another layer of uncertainty in the IRF estimation. Second, and more importantly, we confirm that the responses of both GDP and TFP exhibit a very persistent second hump that is still strongly significant at long horizons.

An additional strategy to assess the *joint* impact of all priors on the posterior distributions is developed by Müller (2012). This method is specifically designed to assess the relative importance of both the priors and the likelihood on the posterior estimates and can be computed for nonlinear transformations of the underlying parameters, such as impulse responses and output multipliers. More specifically, we calculate the measures of Prior Sensitivity (PS_γ) and Prior Informativeness (PI_γ) proposed by Müller (2012). The first metric approximates the largest change of the posterior mean that can be induced by changing the prior mean by the multivariate analog of one prior standard deviation. The second metric, which is contained in the interval $[0, 1]$, summarizes the relative amount of prior information in the posterior distributions. Following the analysis in Müller (2012), we also report the statistic R^2_γ, which is a measure of goodness of fit in a linear regression of the impulse response values on the underlying parameters in the posteriors and priors. This statistic is a useful complement to (PS_γ) and (PI_γ) because it measures the

[16] In this step, we follow Giannone, Lenza, and Primiceri (2015) and allow for both the "Single Unit Root" and "Sum of Coefficients" priors to be present simultaneously. Therefore, we are integrating over four hyperparameters: $\lambda, \alpha, \theta, \mu$. In the Supplemental Appendix, we report prior and posterior distributions of these four hyperparameters.

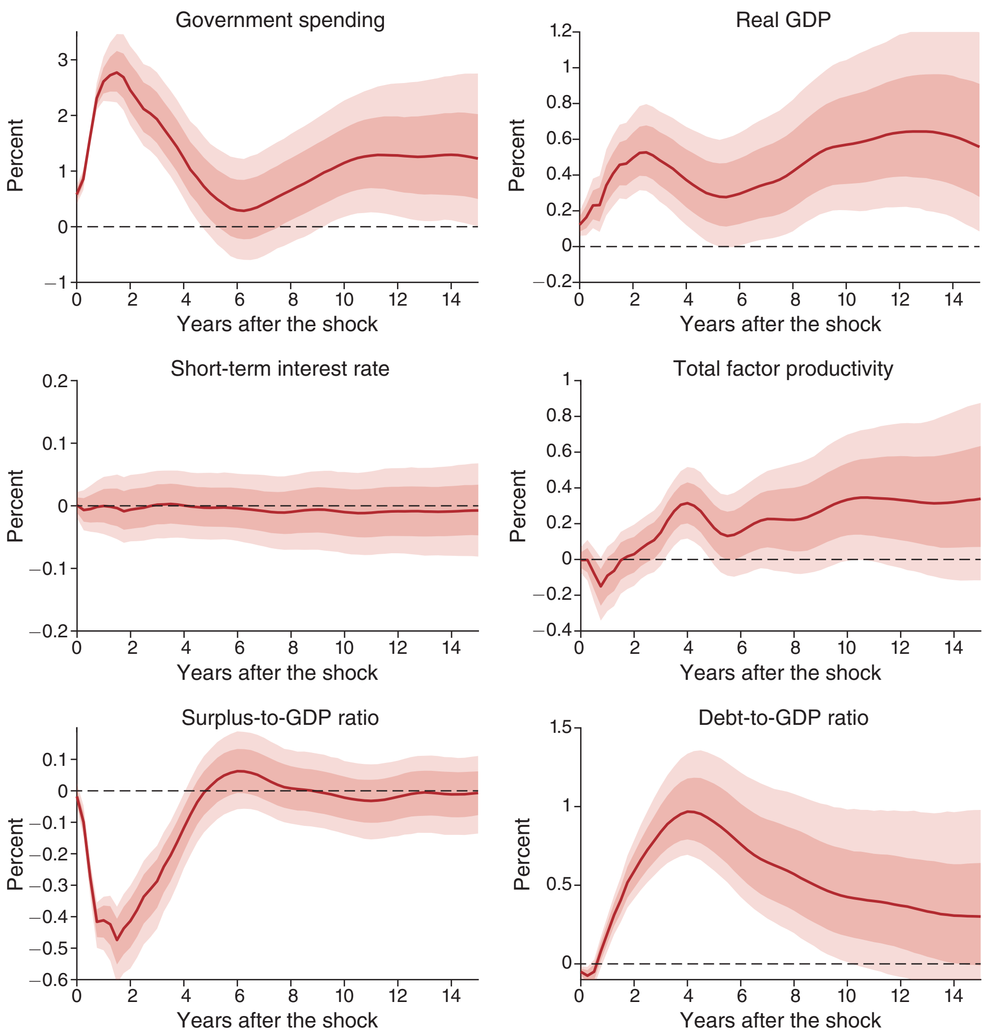


Figure 5. Impulse Responses Using Hierarchical Priors

Notes: The impulse responses are based on an estimated VAR with the same specification of the baseline figure, but where the prior hyperparameters are estimated using the hierarchical MCMC method described in Giannone, Lenza, and Primiceri (2015). The size of the shock is normalized such as to increase government spending by 1 percent of GDP over the first year after the shock, based on a median G/Y ratio of 19 percent. The darker (lighter) shaded areas represent the central 68 percent (90percent) HPD intervals. The darker solid lines are the median estimates. Results are based on 5,000 posterior draws.

validity of approximating the nonlinear transformation of the parameters with a linear function in the Prior Informativeness calculations.

In panel A of Table 1, we report the findings for the GDP impulse responses, whereas in panel B we record those for the output multiplier. Each row represents a different forecast horizon, from 4 up to 60 quarters, whereas the columns refer to the posterior mean, the posterior standard deviation, and the three statistics discussed above, PS_γ, PI_γ, and R^2_γ. In both panels, the measure of prior sensitivity increases with the horizon but always takes very low values. For instance, looking at the GDP

Table 1—Prior Sensitivity and Prior Informativeness Analysis

Horizon (h)	μ_π	σ_π	PS_γ	PI_γ	R^2_γ
Panel A. GDP impulse response					
4	0.29	0.08	0.04	0.00	1.06
8	0.40	0.09	0.04	0.00	1.08
12	0.36	0.09	0.05	0.00	1.04
20	0.16	0.10	0.05	0.00	0.98
40	0.44	0.13	0.06	0.00	1.01
60	0.31	0.18	0.08	0.00	0.96
Panel B. Output multiplier					
4	0.62	0.18	0.08	0.00	1.05
8	0.74	0.16	0.07	0.00	1.09
12	0.87	0.16	0.08	0.00	1.12
20	0.99	0.21	0.10	0.00	1.08
40	1.51	0.34	0.15	0.00	1.04
60	1.77	0.66	0.27	0.00	0.74

Notes: μ_π is the posterior mean of the object of interest; σ_π its standard deviation; PS_γ is the Prior Sensitivity, which approximates the largest change of the posterior mean that can be induced by changing the prior mean by the multivariate analog of 1 prior standard deviation; and $PI_\gamma \in [0,1]$ is the prior informativeness, which summarizes the relative amount of prior information in the posterior. R^2_γ is the R^2 in a linear regression of the impulse response value on the underlying parameters in the posterior and prior.

impulse response at the 60-quarter horizon, a change in the prior mean would induce a change in the posterior that is less than half of the posterior standard deviation. Consistent with this, we find values for the prior informativeness statistics that are essentially zero at all horizons. The results for the output multiplier paint a similar picture. Taken together with the results reported previously, we confirm our assessment that our key results are driven by features of the data rather than features of the prior distributions.

C. *Monte Carlo Analysis*

In this section, we evaluate the ability of our BVAR (60) to draw accurate inference about the long run. Our goal is to establish whether our richly parameterized model has any tendency to spuriously detect long-lasting effects when these are actually *not* present in the data-generating process (DGP). We present two exercises. In our first DGP, the artificial data are fully independently and identically distributed across time:

$$\mathbf{y}_t \sim \mathcal{N}\left(\mathbf{0}_{n\times 1}, \bar{\Sigma}\right) \quad \text{for} \quad t = 1, \dots, T. \tag{6}$$

We set the covariance matrix, $\bar{\Sigma}$, to the posterior mean of our baseline estimates, so as to preserve the contemporaneous correlation structure present in US data. Given this, the theoretical impulse responses in our first DGP have—on impact—the same magnitude as in our baseline results but decay immediately to zero (and remain there) after the first quarter.

In a second, more empirically relevant and more challenging exercise, we consider a DGP in which (i) some variables display persistence at business cycle frequencies; (ii) GDP, government spending, and TFP share a unit root driven by a single TFP shock, ε_t^{TFP}; and (iii) GDP and government spending are also subject to persistent but ultimately transitory shocks. More specifically, we postulate the following processes:

$$m_t = \varepsilon_t^{news}$$

$$TFP_t = TFP_{t-1} + \varepsilon_t^{TFP}$$

$$\tilde{G}_t = 0.1\, m_{t-1} + 1.7\, \tilde{G}_{t-1} - 0.73\, \tilde{G}_{t-2} + \varepsilon_t^{G}$$

$$\tilde{Y}_t = 1.3\, \tilde{Y}_{t-1} - 0.4\, \tilde{Y}_{t-2} + \varepsilon_t^{Y}$$

$$G_t = TFP_t + \tilde{G}_t$$

$$Y_t = TFP_t + \tilde{G}_t + \tilde{Y}_t,$$

where the process for military news, m, is driven by the news shock ε_t^{news}; $\tilde{G}$ is the deviation of government spending from the TFP trend; $\tilde{Y}_t$ refers to the output deviations from the TFP trend; G_t is total government spending; and Y_t is output. The shocks $\left\{\varepsilon_t^{news}, \varepsilon_t^{TFP}, \varepsilon_t^{G}, \varepsilon_t^{Y}\right\}$ are i.i.d. and uncorrelated to each other. The parameters of the process above are calibrated so that the impact of the military shock matches the first peak, as well as the share of the variance, observed for government spending in our baseline estimates. This DGP displays the following properties. First, a shock to TFP induces a single unit root that is common to TFP, output, and government spending; second, a military news shock produces a hump-shaped response in government spending and, insofar as spending forms part of output, a similar response in the latter, with no amplification. In other words, the fiscal multiplier is equal to 1, and government spending has zero long-run effect on output. Third, there are confounding shocks that affect the evolution of both government spending and output. The key question is whether in this setting the confounding unit-root shocks contaminate the estimation of the impulse responses and lead to the detection of spuriously very persistent effects.

In both Monte Carlo analyses, we use a panel of $n = 7$ variables and $T = 504$ observations, which is identical to our baseline sample. Moreover, we always use the same baseline priors: GDP, TFP, and government spending are centered around a unit root; the other variables are centered around an AR(1) coefficient with persistence parameter equal to 0.9; the tightness of the hyperparameters is set to the same values used in Section IIA. In Table 2, we record the results of both exercises. Across the rows, we report estimates about the GDP response to the military news shock at various horizons. The first column reports the median difference between the point estimate and the true IRF across MCMC experiments; the second column reports the median length of the 90 percent confidence interval; and the third column represents the coverage probability, that is, the fraction of MCMC draws for which the true value lies inside of the estimated 90 percent confidence intervals.

For the i.i.d. data-generating process in panel A, the point estimates are centered around the true value at all horizons, with narrow confidence intervals. The

Table 2—Monte Carlo Analysis

Horizon (h)	Median bias	Median length	Coverage prob.
Panel A. Fully i.i.d. data-generating process			
4	0.00	1.33	0.91
8	0.01	1.02	0.97
12	0.00	0.77	0.97
20	0.00	0.53	1.00
30	0.00	0.37	1.00
40	0.00	0.28	1.00
60	0.00	0.19	1.00
Panel B. Single unit root data-generating process			
4	−0.03	0.06	0.52
8	−0.06	0.13	0.49
12	−0.06	0.16	0.61
20	−0.05	0.17	0.75
30	−0.03	0.17	0.71
40	−0.02	0.17	0.78
60	−0.01	0.18	0.87

Notes: Median bias of point estimate, median length, and coverage probability of nominal 90 percent confidence intervals at different horizons. Data-generating processes as described in the main text. $T = 504$, i.i.d. standard normal innovations. 100 Monte Carlo repetitions; 1,000 Gibbs sampler iterations.

coverage probability is close to the nominal level at the short horizons but rises above 90 percent at medium to long horizons, indicating confidence intervals that—if anything—are too conservative in this setting. Based on this, it seems unlikely that our estimation procedure would find long-lasting effects if the data were actually characterized by no persistence at all.

As for the more challenging data-generating process of panel B, we find that while there is some evidence of a possible undercoverage at short horizons, there is simultaneously evidence of a small *downward* bias, and at intermediate and long horizons—which are the main focus of our analysis—the impulse responses are centered around the true value, and the coverage probabilities are only slightly below the nominal level of 90 percent. We conclude that it is highly unlikely that our finding of positive and very persistent effects at frequencies beyond the business cycle is driven by either our prior specification or by the presence of confounding unit roots in the data that are common to both GDP and government spending.

D. *Frequentist Local Projections*

A further way to assess whether our results are driven by features of the Bayesian VAR is to use frequentist local projections, without any shrinkage. Making the two methods comparable is challenging because the large number of lags used as controls in the VAR will quickly erode the degrees of freedom in the LPs and thus lead to very imprecise estimates when no shrinkage is used. The trade-off between bias and variance in LPs and the attractiveness of shrinkage methods are discussed in detail by Li, Plagborg-Møller, and Wolf (2021a). Accordingly, we reduce the number of lags to 20, except for the military spending news and the outcome variable, where we keep 60 lags. The LP estimates are reported in Figure 6. Standard errors are adjusted for heteroskedasticity and autocorrelation (HAC). The IRFs produced by the LPs

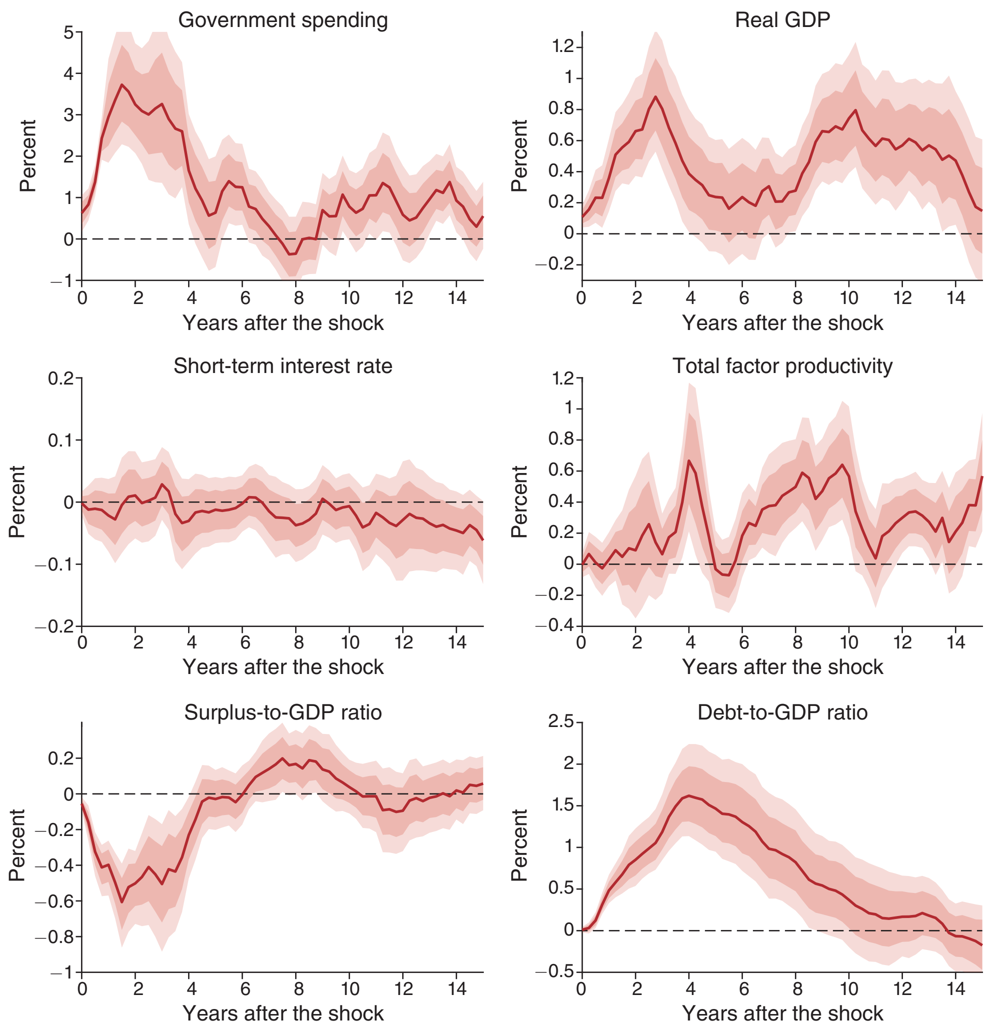


Figure 6. IRFs to Military News Shock from Frequentist Local Projections

Notes: In each panel, the impulse responses are frequentist estimates of local projections with 60 lags of the military spending news and the outcome variable as well as 20 lags of all remaining variables in our baseline 7-variable dataset. Government spending, GDP, and TFP enter the VAR in log-levels. The size of the shock is normalized such as to increase government spending by 1 percent of GDP over the first year after the shock, based on a median G/Y ratio of 19 percent. The darker (lighter) shaded areas represent the central 68 percent (90 percent) frequentist confidence intervals. The darker solid lines are the mean estimates.

appear more jagged than their BVAR counterparts but confirm the very significant and persistent responses of output and TFP at horizons beyond the business cycle. In the Supplemental Appendix, we present the results of an alternative strategy to solve the curse of dimensionality.[17] This involves collapsing the $60 \times 7 = 420$ original controls into $k = 43$ principal components that explain the bulk of their variance.[18]

[17] We thank an anonymous referee for suggesting the two approaches used in this subsection.
[18] We select the optimal number of principal components according to the criteria proposed by Bai (2004).

This exercise points to qualitatively similar medium-run effects on output and productivity following a military spending news shock.

IV. Inspecting the Mechanism

In the previous sections, we have reported extensive evidence of robustly significant and highly persistent responses of output and productivity to a change in military spending. To shed light on the transmission mechanism, in this section we look at the effects of government spending shocks on private sector outcomes and public spending categories.

To mitigate the curse of dimensionality, in each specification, we augment our baseline VAR (60) with at most two variables at a time, which also enter the model in log-levels. For the private sector models, we consider the following four pairs of additions: (i) labor productivity and hours worked (which substitute for GDP), (ii) unadjusted total factor productivity and patents, (iii) private consumption and private investment, (iv) exports and imports. For the public sector specifications, we add in turn (v) public consumption expenditure, (vi) public investment in E&S, and (vii) public expenditure in R&D.

A. *Private Sector*

In Figure 7, we report the posterior credible sets around the responses to a government spending shock, based on the four extended VARs (60) for the private sector described above. The first row focuses on the specification with labor productivity and hours worked, the second row refers to the model with adjusted TFP and innovation (as measured by patents), the third row reports the estimates for consumption and investment changes, and the fourth row summarizes the responses of exports and imports. Several results emerge from Figure 7. First, after a short-lived decline, labor productivity experiences a sustained increase, which peaks significantly at the end of the forecast horizon. In contrast, hours worked rise on impact and peak in their first year (consistent with the temporary productivity drop) before recording small and insignificant changes.

The second row reveals that the effects of government spending shocks on labor productivity are mirrored by the dynamics of unadjusted TFP. Given the low statistical power of empirical time series model to distinguish between permanent and very persistent effects, we stress once more that our preferred interpretation of our evidence is that the effects of government spending on output and productivity are large and significant well beyond the five years forecast horizon typically considered in empirical business cycle analyses. In contrast, the right-hand panel of the second row makes clear that patents are crowded out in the first few years after the shock; their response, however, turns positive and significant in the medium to long term, consistent with the findings in Diebolt and Pellier (2020) that infrequent, large shocks, such as wars, account for the largest pushes in innovation and the very process of economic growth in the United States over the last century. Both IRFs in the second row of Figure 7 are also consistent with the micro evidence on patents in Comin and Mestieri (2014) and Pezzoni, Veugelers, and Visentin (2022), who estimate an average adoption lag between 6 and 17 years in the rate of technological diffusion over the post-WWII period.

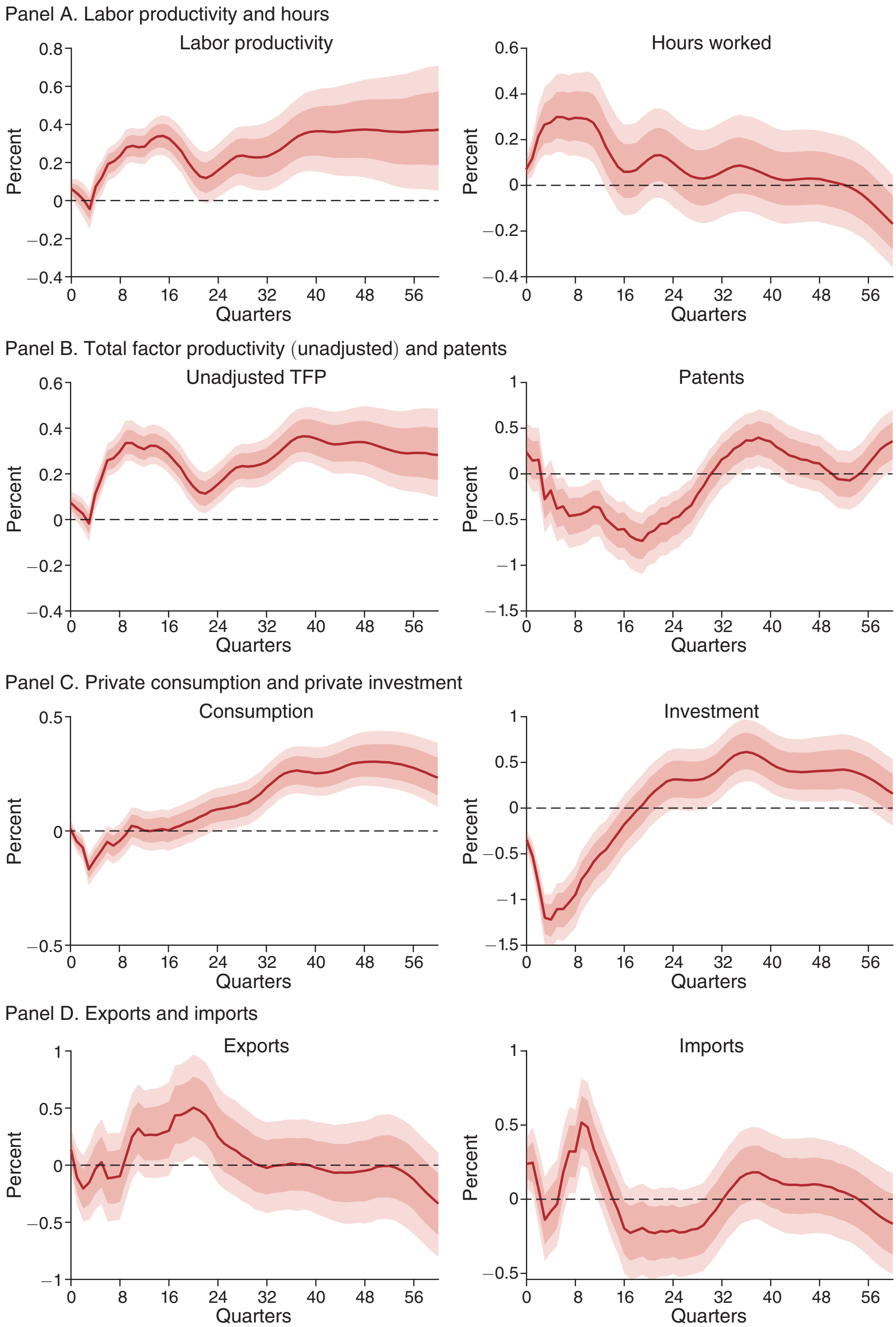


Figure 7. Effects of Military Shocks on Private Sector Outcomes

Notes: The impulse responses are based on an estimated VAR with 60 lags adding to the baseline specification the series in each panel. Military spending news is ordered first in the Cholesky factorization. The size of the shock is normalized such as to increase total government spending by 1 percent of GDP over the first year after the shock. The darker (lighter) shaded area represents the central 68 percent (90 percent) HPD band. The darker solid line stands for the median estimates. Results are based on 5,000 posterior draws.

In the third row of Figure 7, a public spending hike crowds out private consumption and private investment in the short run, as reported also by Ramey (2011b). Five years after the shock, however, both expenditures go up significantly, peaking at horizons of about 9 to 12 years. The magnitude of the investment response is larger than the size of the consumption effect, possibly reflecting its more volatile nature and smaller GDP share. Finally, military spending has a significant short-term impact on imports and a delayed effect on exports, which imply a significant trade surplus between two and six years after the shock.

In summary, government spending causes a short-lived rise in hours worked; a temporary crowding out of innovation, private investment, and consumption; and a delayed hump in net exports. In the medium to long run, however, investment and innovation experience significant and sustained increases, which feed into large and very persistent effects on labor and total factor productivity as well as consumption.

B. *Public Sector*

The findings in the previous section are consistent with an important role played by productivity, innovation, and private investment in shaping the responses of output to a government spending shock at long horizons. In this section, we ask whether the particular composition of public spending triggered by a defense budget increase may also have a significant contribution. To this end, we run three separate model specifications in which we augment the baseline VAR (60) of Section II with our newly reconstructed historical time series of public consumption, government investment, and public R&D, respectively.

The results of these public sector–augmented VARs (60) are reported in Figure 8. The first row depicts the response of the *log-level* of each category to the military spending news, while the second row reports the response of each category as a *share* of total government spending. The top panels reveal three main findings. First, the responses of these three public spending categories are highly correlated: Military spending triggers a joint increase in public consumption, investment, and R&D.[19] Second, public investment is the category that responds most in the short run. Third, government R&D expenditure is the only component that displays a large and persistent response.

To appreciate the relative contribution of each category, in the second row of Figure 8, we look at the responses of public consumption, investment, and R&D as *shares* of total government spending. Given the data are in logarithms, these are computed as the difference between the impulse response of each spending category and the impulse response of total government spending at each horizon.[20] Three results stand out from this exercise. First, following a military spending shock, there

[19] We interpret this finding as a cautionary note *against* counterfactual exercises that try to isolate the effects of a specific public spending category by setting to zero at all times the responses of all other components of the government budget. In the context of military spending (and possibly also of other large public programs), this "counterfactual" mix is actually well outside the distribution of historical combinations of public spending components.

[20] Over our long sample, consumption, investment in E&S, and R&D expenditure account, on average, for about 77 percent, 20 percent, and 3 percent of total government spending. During the post-WWII period, the average share of public R&D has increased to around 5 percent, offset almost entirely by a decline in the share of public investment.

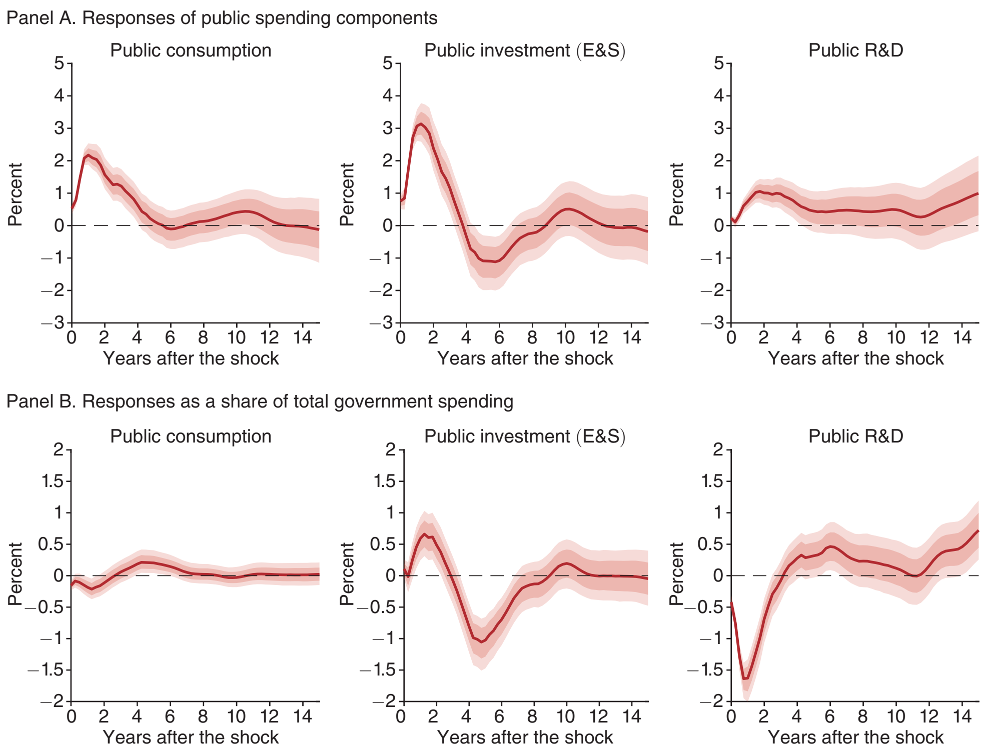


Figure 8. Effects of Military Shocks on Public Spending Components

Notes: The impulse responses are based on an estimated VAR with 60 lags of military spending news, real government spending per capita, real GDP per capita, utilization-adjusted TFP, short-term interest rate, fiscal deficit to GDP ratio, and government debt to GDP ratio. In each of the columns, real public consumption per capita, real public non-R&D investment per capita, and real public R&D expenditure per capita, respectively, are added in turn to the VAR. Each government spending category, total government spending and output enter the VAR in log-levels. Military spending news is ordered first in the Cholesky factorization. The size of the shock is normalized such as to increase total government spending by 1 percent of GDP over the first year after the shock. The top (bottom) panel refers to the response of each public spending category in log-level (as share of total government spending). The darker (lighter) shaded area represents the central 68 percent (90 percent) HPD band. The darker solid line stands for the median estimates. Results are based on 5,000 posterior draws.

are only small movements in the consumption share, except for two small drops at the start and at the end of the forecast horizon. Second, in the short run, the composition of public spending shifts significantly toward investment and away from R&D. Third, in sharp contrast, at intermediate and long horizons, the share of public R&D records a significant increase, which is offset by a decline in the public investment share and, to a lesser extent, in the consumption share.

One interpretation of the responses of the different public spending categories is that public investment in E&S plays a far more important role in explaining the effects of government spending at *short* horizons, whereas public R&D expenditure plays a far more important role in accounting for the effects of government spending at *long* horizons. In the next section, we explore this result further by digging deeper into the role that R&D plays in driving the effects on long-run output.

V. What Drives the Long-Run Effects of Government Spending?

In the previous section, we have shown that military spending has very significant and persistent effects: (i) on public R&D (but not on public consumption or investment) and (ii) on productivity and output. In this section, we ask whether changes in government R&D spending are a main driver of the significant output response at long horizons. Our analysis proceeds in three steps. We start by documenting a significant reduced-form correlation between public R&D and output at low frequencies, based on either our BVAR (60) or the factor model proposed by Müller and Watson (2020). Then, we put forward a simple but intuitive strategy to identify public R&D spending shocks and show that these cause large and very persistent effects on GDP and TFP. Finally, we report the IRFs to public consumption and public investment shocks and find evidence of virtually no response of output and TFP at intermediate and long horizons.

A. *The Low-Frequency Covariability of Public R&D and GDP*

In this section, we look at the reduced-form, low-frequency comovement between government R&D and GDP. We do so by relying on methods recently developed by Müller and Watson (2020), which are specifically designed for conducting inference on covariability among economic time series at very long horizons. Following Müller and Watson (2020), we specify a low-frequency factor model for the growth rates of government R&D, real GDP, and TFP. The inference is based on only a small number of periodic functions that capture the low-frequency properties of the data. More specifically, we focus on frequencies lower than 15 years, consistent with the lag length of our baseline BVAR as well as the forecast horizon that we have used in the IRF analyses throughout the paper. The results of the factor model in Müller and Watson (2020), and how it compares to the BVAR (60), are summarized in the Supplemental Appendix, which also reports a description of the low-frequency factor model. Despite the important methodological differences, both models estimate significant low-frequency correlations between the three variables and produce very similar unconditional forecasts at 25- and 50-year horizons.

B. *Identifying Public R&D Shocks*

The ideal experiment to isolate exogenous movements in government R&D would consist of "shocking" public R&D while keeping fixed both public consumption and public investment. But this very specific policy mix has virtually never happened in our long historical sample, as government spending typically involves a simultaneous expansion in all three categories.[21] Insofar as the correlation is not perfect, however, we can use a statistical approach to tease out the effects of each public spending category on the US economy.

[21] Interestingly, the evidence in Figure 8 reveals that military spending comes close to an ideal (long-run) experiment, as it is associated with a significantly long-lasting response of public R&D but very small and short-lived responses of public consumption and investment. However, the short-run dynamics are very different.

Our starting point is to note that, historically, the major shifts in public R&D spending have been unrelated to business cycle conditions. In the Supplemental Appendix, we discuss the narrative evidence around large public R&D programs and argue that, over our long sample, these have been, in fact, motivated by military rivalries (with Germany until WWII and the Soviet Union afterward), scientific progress, and ideological priorities of the different administrations, rather than by an endogenous response to the state of the US economy.

In addition, the timing and implementation lags associated with large public R&D programs extend well beyond business cycle frequencies or the terms of office of the different administrations. These considerations suggest that, after controlling for the lags of other macro variables, innovations to public R&D expenditure may be regarded as good as exogenous to current or prospective economic conditions, in the spirit of the short-run restrictions on government spending proposed by Blanchard and Perotti (2002) or the narrative identification for income tax changes pioneered by Romer and Romer (2010).

In practice, we drop military spending news from the model and add public R&D, patents, and private investment. We then identify exogenous changes in public R&D by searching for the shock that explains the maximum share of the public R&D innovation variance over the first year, following Uhlig (2004).[22] We focus on the first year, rather than the first quarter, because much of our historical data have been interpolated from annual series and the interpolation method might spuriously affect some of the high-frequencies correlations.

Before presenting the impulse response analysis, we find it useful to verify whether our newly identified shock can match the historical evolution of large federal R&D programs, as discussed in the Supplemental Appendix. To this end, in panel A of Figure 9, we present the historical decomposition of public R&D around three key historical events: (i) the Manhattan Project, from its establishment in 1941 to its dissolution in 1946 with the foundation of the Atomic Energy Commission; (ii) the creation of DARPA in 1958 and the moon landing project from 1961 to 1969; and (iii) Reagan's Strategic Defense Initiative from 1983 to 1987. In each subpanel, the solid black line represents the historical increase in public R&D, while the dotted blue line, and associated 68 percent posterior bands, refers to the part explained by our public R&D shock. In all cases, the movements in government R&D attributed to the shock align very closely with the actual increases around the three events. We interpret this as suggestive evidence that our shock captures the exogenous nature of military or ideologically driven surges in public R&D.[23]

In panel B of Figure 9, we report the time series of the identified public R&D shock, together with 68 percent posterior bands. The shock is plotted as an eight quarter moving average. Two findings are worth noting. First, there are clusters of positive shocks around the three major public R&D programs (vertical dashed

[22] The "max-share" method generalizes to any desired frequency the well-known Cholesky decomposition. The latter imposes the far more restrictive restriction that the identified shock explains the entirety of the variance of the variable of interest on impact. The "max-share" method has been shown to be more robust than the Cholesky factorization in a variety of empirical settings (see, e.g., Kurmann and Otrok 2013; Francis et al. 2014).

[23] In the Supplemental Appendix, we show that—in sharp contrast to the results in this section—the military spending shocks cannot explain the lion's share of movements in public R&D expenditure around these three key historical events.

lines). Similarly, a cluster of negative shocks is visible around the wind-down of the Apollo project. Second, the timing of these programs does not always coincide with major wars (shaded areas). For instance, while WWI and WWII led to large increases in defense-related R&D, the Korean war did not. In other words, the R&D shock seems distinct from the military news shock. Indeed, the sample correlation of the posterior mean of the two shocks is only 0.17.

In Figure 10, we report the IRFs to the public R&D shock. In keeping with previous charts, the shock is scaled so as to increase total government spending by 1 percent of GDP over the first year. At short horizons, the increase in output is much more muted than for the military spending shock and does not display any hump shape. At long horizons, however, the size and persistence of the effects on output and TFP become much larger, with a peak toward the end of the forecast horizon. Interestingly, using a very different identification strategy based on a narrative classification of R&D appropriations over a post-WWII sample, Fieldhouse and Mertens (2023) report long-lasting effects of a public R&D shock on US productivity and GDP trend that are similar, both in size and duration, to the estimates in Figure 10. Finally, the responses of private investment and patents display dynamics that are qualitatively in line with those produced by the military spending shock in Figure 1. For both variables, however, the public R&D shock causes a smaller short-run crowding-out effect, which is no longer statistically significant for patents.

In summary, our identified public R&D shock aligns very well with the narrative account around large public R&D programs in the economic history of the United States. Furthermore, we find that the effects on output, productivity, private investment, and innovation generated by an exogenous increase in government R&D are qualitatively similar to, if not more persistent than, those triggered by military spending, despite a relatively modest correlation between the two identified shocks.[24]

C. *The Role of Public Consumption and Public Investment*

In the previous section, we have identified the effects of public R&D on the economy by searching for the shock that explains most of the public R&D variance during the first year after the shock. For sake of comparability, in this section we adopt an identical strategy for the other two components of government spending and isolate the innovations to public consumption and public investment, respectively, that maximize the share of the forecast error variance of each spending component at the one-year horizon.

It is worth noting, however, that both shocks are in fact associated with significant contemporaneous movements in public R&D, which makes it hard to interpret them as "pure" innovations (i.e., everything else equal) to public consumption and investment. On the other hand, each innovation brings about movements in public R&D of different sizes, and therefore, we can exploit this variation to assess whether the strength of the output responses is correlated with the relative strength or "intensity" of the changes in public R&D.

[24] In the Supplemental Appendix, we further show that the estimated impulse responses for GDP and TFP in Figure 10 are very similar to those obtained by identifying and estimating the effects of a public R&D shock over a post-WWII sample that is characterized by no major war involving the United States.

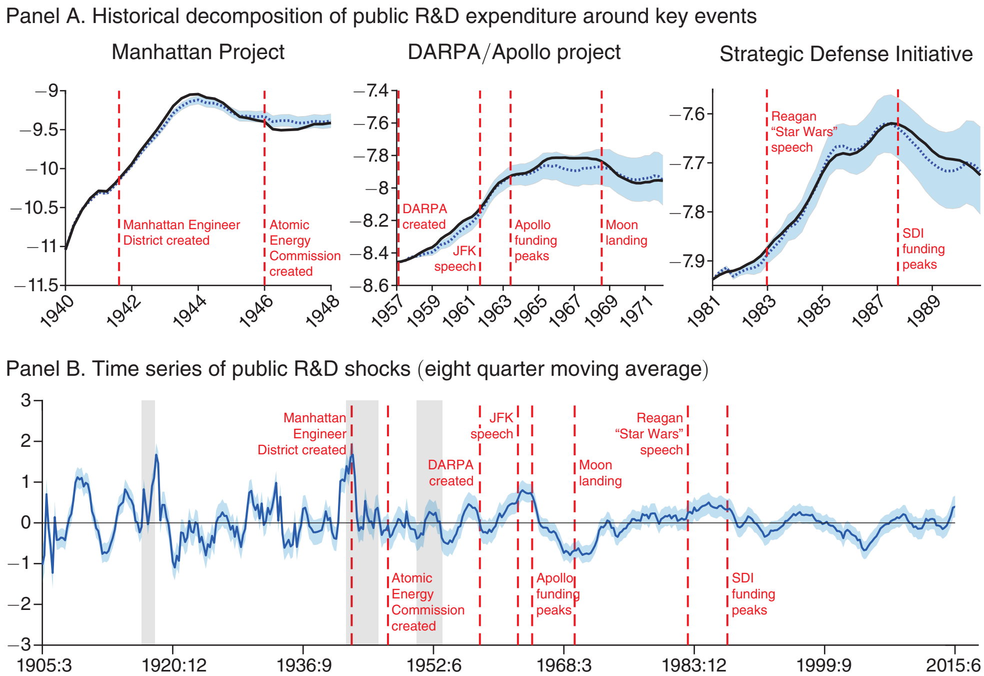


Figure 9. Historical Analysis of Public R&D and Public R&D Shocks

Notes: Panel A plots the historical decomposition of the public R&D series around three key historical events: (i) the Manhattan Project, (ii) DARPA and the Apollo program, and (iii) the Strategic Defense Initiative. In each of the subpanels, the solid black line is the historical increase in real per capita R&D spending by the government. The dotted blue line, and associated 68 percent posterior bands, shows the part of the increase in R&D that can be explained by the effects of the exogenous public R&D shock identified using the max-share method at the one-year forecast horizon. Panel B plots the history of identified public R&D together with 68 percent posterior bands. To facilitate visualization, the shock is plotted as an eight-quarter moving average. Shaded areas represent major wars.

In panel A of Figure 11, we report the output responses to shocks that maximize the one-year-ahead error variance of public consumption, public investment, and public R&D, respectively.[25] Across all specifications, the shocks are normalized such that total government spending moves by 1 percent of GDP over the first year; hence, the three columns can also be thought of as varying the intensity of each spending category. The main finding is that the "consumption-intensive" shock leads to a smaller output response than the "investment-intensive" shock, which in turn triggers a smaller response than the "R&D-intensive" shock.

To explore these results further, in panel B, we look at the effects of each shock on public R&D as a share of total government spending. The shock to public consumption leads to a drop in the R&D share in the short run and a muted response thereafter. This is associated with modest effects on output at long horizons in panel A. The shock to public investment in the middle column also leads to a short-run decline in the R&D share; this is, however, quickly reversed and then replaced by a persistent

[25] The chart in the top-right panel of Figure 11 is therefore a repetition of the top-left panel in Figure 10.

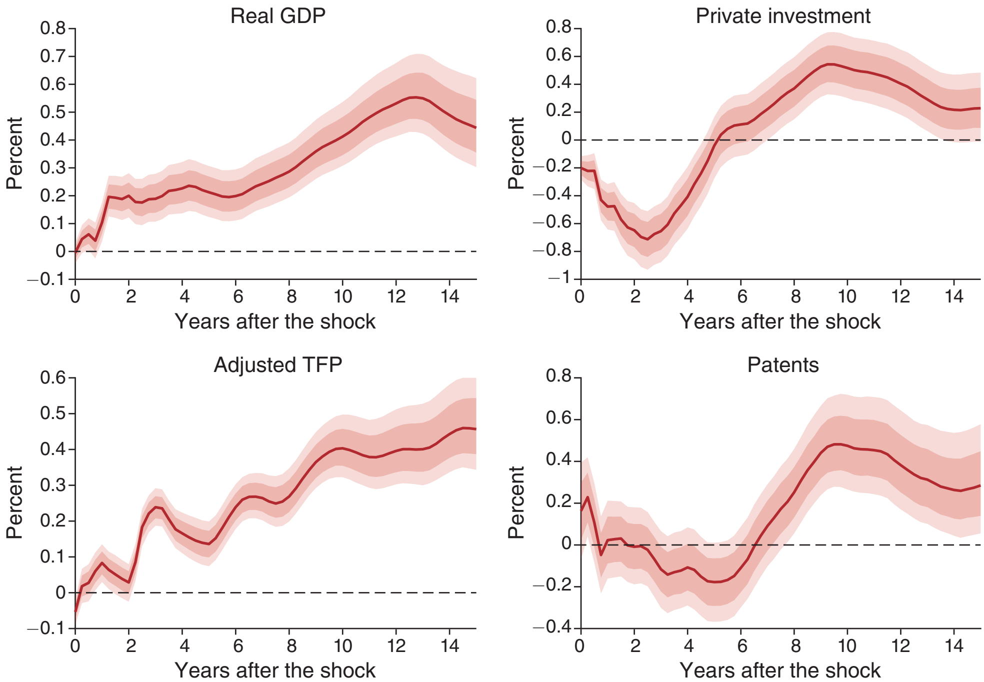


Figure 10. Impulse Responses to Public R&D Shock

Notes: The impulse responses are based on an estimated VAR with 60 lags of real public R&D per capita, real total government spending per capita, real GDP per capita, utilization-adjusted TFP, short-term interest rate, fiscal deficit to GDP ratio, government debt to GDP ratio, real private investment per capita, total factor productivity, and patents. Public R&D, total government spending, GDP, and TFP enter the VAR in log-levels. The public R&D shock is identified using the max-share method at the one-year forecast horizon. The size of the shock is normalized such as to increase government spending by 1 percent of GDP over the first year after the shock, based on a median G/Y ratio of 19 percent. The darker (lighter) shaded areas represent the central 68 percent (90 percent) HPD intervals. The darker solid lines are the median estimates. Results are based on 5,000 posterior draws.

increase, which is mirrored by the output response in the top row of Figure 11.[26] On the other hand, the R&D shock is characterized by both the largest R&D share response *and* the largest and most persistent output response.

In summary, military spending shifts the composition of public spending toward R&D. A shock that raises the relative intensity of government R&D leads to large and persistent responses of investment, productivity, innovation, and output. The latter is far larger and more persistent than the output responses to either more "public consumption–intensive" or more "public investment–intensive" shocks. We interpret this as suggestive evidence that public R&D is a key driver of the effects of government spending on output beyond business cycle frequencies documented in this paper.

[26]This is likely to reflect the patterns of military spending ramp-up, which, as discussed around Figure 8, lead to large short-run responses of investment and a longer-run increase in the share of R&D. Unsurprisingly, the output response to the public investment shock is more similar to the output response to the military spending shock.

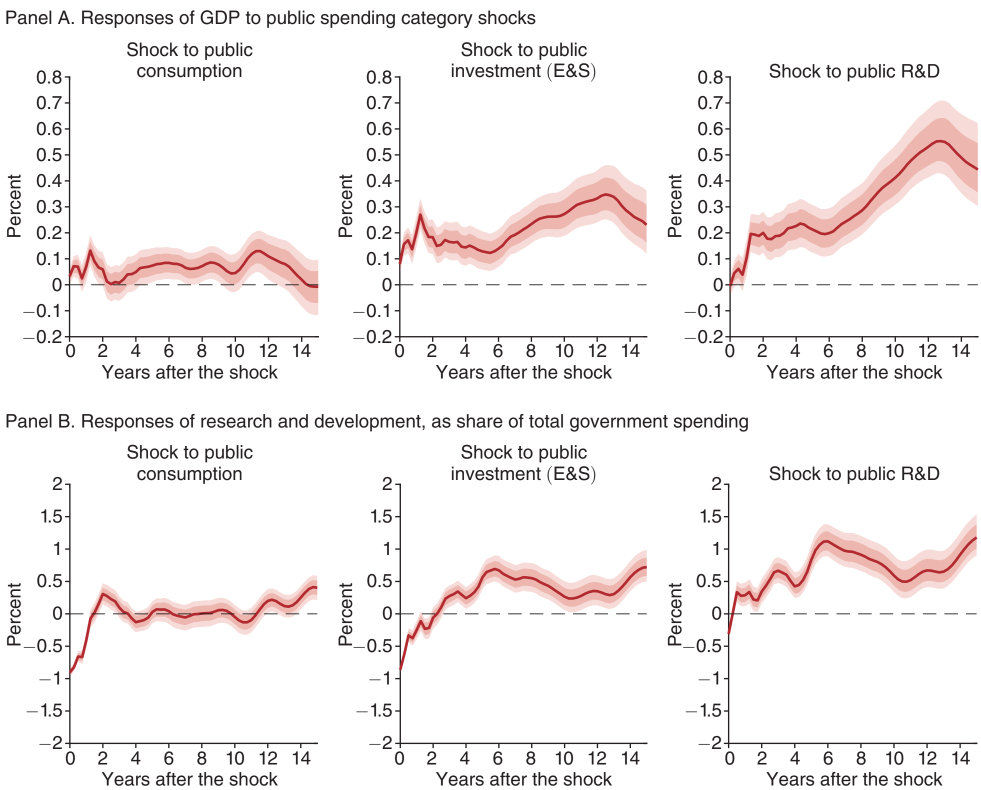


Figure 11. Impulse Response to Public Spending Category Shocks

Notes: The impulse responses are based on an estimated VAR with 60 lags of real government spending per capita, real GDP per capita, utilization-adjusted TFP, short-term interest rate, fiscal deficit to GDP ratio, and government debt to GDP ratio. In each of the columns, real public consumption per capita, real public non-R&D investment per capita, and public R&D expenditure per capita, respectively, are added in turn to the VAR. Each public spending category, total government spending, GDP, and adjusted TFP enter the VAR in log-levels. The shock to each public spending category is identified using the max-share method at the one-year forecast horizon for that category. The size of the shock is normalized such as to increase total government spending by 1 percent of GDP over the first year after the shock. The top (bottom) panel refers to the response of GDP (public R&D as share of total government spending) to shocks to each public spending category. The darker (lighter) shaded area represents the central 68 percent (90 percent) HPD band. The darker solid line stands for the median estimates. Results are based on 5,000 posterior draws.

VI. Conclusions

What are the long-term effects of government spending? Despite the resurgence in fiscal research spurred by the financial crisis of 2007–2009 and the policy debate triggered by the global pandemic of 2020–2022, this question seems to have so far eluded empirical research. We use 125 years of US quarterly data—including newly constructed series of public spending by main categories—and time series models with long lags to shed light on this question. We argue that the combination of historical data, a generous lag length selection, and Bayesian shrinkage makes our framework well suited to draw inference about very persistent dynamics and diffusion patterns, while retaining the ability to look also at the short run.

We uncover four main regularities. First, fiscal policy can stimulate economic activities persistently when it tilts the share of public spending toward R&D, as it does, for instance, during military conflicts. However, we also find that an exogenous increase in public R&D expenditure can have very persistent effects on output and productivity even when it is not systematically associated with war spending. Second, in contrast, government investment has shorter-lived effects, whereas the impact of public consumption on output is modest at most horizons. Third, while government spending crowds out innovation, private investment, and private consumption in the short run, it crowds them in over the medium term, feeding into a very sustained increase in productivity. As a result, the government spending multiplier on output raises above one at longer horizons. Our evidence uncovers a novel mechanism through which fiscal policy can stimulate the economy and productivity, even beyond the business cycle.

Our analysis exploits low-frequency variations and comovements among total government spending, military purchases, public R&D, productivity, and GDP over a historical sample of 125 years, with the goal of identifying the effects of government spending on the US economy at horizons of up to 15 years. The significant advantage of using a long historical sample to identify low-frequency covariability comes, however, at the cost of facing possible subsample instability induced by the fact that economic policy priorities, the structure of the economy, and policy responses to business cycle conditions have varied greatly across the many administrations that spanned the last century. Balancing the trade-off between maximizing low-frequency variation that could help identify long-run effects and minimizing subsample instability that might affect long-run inference is a challenge that we leave for future research.

Finally, the government spending shocks proposed by Ramey (2011b) and Ramey and Zubairy (2018) are based on large and infrequent military outbursts, whereas our R&D spending shock identification essentially exploits the virtually acyclical nature of government R&D payments. Less is known, however, on other forms of public spending. A main challenge is the lack of strong and reliable instruments that may shed light on the effects of the different components as well as of total government spending. Further progress in identifying the effects of public expenditure could be made by taking a more granular approach across different government agencies or policies, as in Cox et al. (2024), and combining it with a careful narrative evaluation of long historical record of congressional documents along the lines of Romer and Romer (2010) or Fieldhouse and Mertens (2023), but spanning a much longer sample. We view this Herculean task as the next milestone in the fiscal policy research agenda.

REFERENCES

Akcigit, Ufuk, John R. Grigsby, Tom Nicholas, and Stefanie Stantcheva. 2022. "Taxation and Innovation in the 20th Century." *Quarterly Journal of Economics* 137 (1): 329–85.

Antolin-Diaz, Juan, and Paolo Surico. 2022. "The Long-Run Effects of Government Spending." CEPR Discussion Paper 17433.

Antolin-Diaz, Juan, and Paolo Surico. 2025. *Data and Code for: "The Long-Run Effects of Government Spending."* Nashville, TN: American Economic Association; distributed by Inter-university Consortium for Political and Social Research, Ann Arbor, MI. https://doi.org/10.3886/E198830V1.

Anzoategui, Diego, Diego Comin, Mark Gertler, and Joseba Martinez. 2019. "Endogenous Technology Adoption and R&D as Sources of Business Cycle Persistence." *American Economic Journal: Macroeconomics* 11 (3): 67–110.

Bai, Jushan. 2004. "Estimating Cross-Section Common Stochastic Trends in Nonstationary Panel Data." *Journal of Econometrics* 122 (1): 137–83.

Barnichon, Regis, and Christian Brownlees. 2019. "Impulse Response Estimation by Smooth Local Projections." *Review of Economics and Statistics* 101 (3): 522–30.

Barro, Robert J., and Charles J. Redlick. 2011. "Macroeconomic Effects from Government Purchases and Taxes." *Quarterly Journal of Economics* 126 (1): 51–102.

Beaudry, Paul, Dana Galizia, and Franck Portier. 2020. "Putting the Cycle Back into Business Cycle Analysis." *American Economic Review* 110 (1): 1–47.

Benigno, Gianluca, and Luca Fornaro. 2018. "Stagnation Traps." *Review of Economic Studies* 85 (3): 1425–70.

Bergeaud, Antonin, Gilbert Cette, and Rémy Lecat. 2016. "Productivity Trends in Advanced Countries between 1890 and 2012." *Review of Income and Wealth* 62 (3): 420–44.

Bianchi, Francesco, Howard Kung, and Gonzalo Morales. 2019. "Growth, Slowdowns, and Recoveries." *Journal of Monetary Economics* 101: 47–63.

Blanchard, Olivier, and Roberto Perotti. 2002. "An Empirical Characterization of the Dynamic Effects of Changes in Government Spending and Taxes on Output." *Quarterly Journal of Economics* 117 (4): 1329–68.

Bureau of the Census. 1949. *Historical Statistics of the United States, 1789–1957*. United States Department of Commerce.

Bush, Vannevar. 1954. *Science, the Endless Frontier*. Princeton University Press.

Cerra, Valerie, Antonio Fatás, and Sweta C. Saxena. 2022. "Hysteresis and Business Cycles." *Journal of Economic Literature* 61 (1): 181–225.

Chodorow-Reich, Gabriel. 2019. "Geographic Cross-Sectional Fiscal Spending Multipliers: What Have We Learned?" *American Economic Journal: Economic Policy* 11 (2): 1–34.

Chow, Gregory C., and An-loh Lin. 1971. "Best Linear Unbiased Interpolation, Distribution, and Extrapolation of Time Series by Related Series." *Review of Economics and Statistics* 53 (4): 372–75.

Cloyne, James, Joseba Martinez, Haroon Mumtaz, and Paolo Surico. 2025. "Corporate Taxes, Innovation and Productivity." Unpublished.

Comin, Diego, and Mark Gertler. 2006. "Medium-Term Business Cycles." *American Economic Review* 96 (3): 523–51.

Comin, Diego, and Martí Mestieri. 2014. "Technology Diffusion: Measurement, Causes, and Consequences." In *Handbook of Economic Growth*, Vol. 2, edited by P. Aghion and S. Durlauf, 565–622. Elsevier.

Cox, Lydia, Gernot J. Müller, Ernesto Pastén, Raphael Schoenle, and Michael Weber. 2024. "Big G." *Journal of Political Economy* 132 (10): 3260–97.

De Lipsis, Vincenzo, Matteo Deleidi, Mariana Mazzucato, and Paolo Agnolucci. 2022. "Macroeconomic Effects of Public R&D." Preprint, SSRN. https://dx.doi.org/10.2139/ssrn.4301178.

Del Negro, Marco, and Frank Schorfheide. 2011. "Bayesian Macroeconometrics." In *Oxford Handbook of Bayesian Econometrics*, edited by John Geweke, Gary Koop, and Herman Van Dijk, 293–389. Oxford University Press.

Deleidi, Matteo, and Mariana Mazzucato. 2021. "Directed Innovation Policies and the Supermultiplier: An Empirical Assessment of Mission-Oriented Policies in the US Economy." *Research Policy* 50 (2): 104151.

Diebolt, Claude, and Karine Pellier. 2020. "Patents in the Long Run: Theory, History and Statistics." In *History and Mathematics: Investigating Past and Future*, Vol. 8, edited by Leonid E. Grinin and Andrey V. Korotayev, 80–119. Uchitel Publishing House.

Doan, Thomas, Robert Litterman, and Christopher Sims. 1984. "Forecasting and Conditional Projection Using Realistic Prior Distributions." *Econometric Reviews* 3 (1): 1–100.

Donaldson, Dave, and Richard Hornbeck. 2016. "Railroads and American Economic Growth: A 'Market Access' Approach." *Quarterly Journal of Economics* 131 (2): 799–858.

Dupree, A. Hunter. 1986. *Science in the Federal Government: A History of Policies and Activities*. Johns Hopkins University Press.

Fernald, John G. 1999. "Roads to Prosperity? Assessing the Link between Public Capital and Productivity." *American Economic Review* 89 (3): 619–38.

Fernald, John G. 2012. "A Quarterly, Utilization-Adjusted Series on Total Factor Productivity." Federal Reserve Bank of San Francisco Working Paper 2012-19.

Fieldhouse, Andrew J., and Karel Mertens. 2023. "The Returns to Government R&D: Evidence from US Appropriations Shocks." Federal Reserve Bank of Dallas Working Paper 2305.

Francis, Neville, Michael T. Owyang, Jennifer E. Roush, and Riccardo Dicecio. 2014. "A Flexible Finite-Horizon Alternative to Long-Run Restrictions with an Application to Technology Shocks." *Review of Economics and Statistics* 96 (4): 638–47.

Friedman, Milton. 1952. "Price, Income, and Monetary Changes in Three Wartime Periods." *American Economic Review* 42 (2): 612–25.

Friedman, Milton, and Anna Jacobson Schwartz. 1963. *A Monetary History of the United States, 1867–1960*. Princeton University Press.

Giannone, Domenico, Michele Lenza, and Giorgio E. Primiceri. 2015. "Prior Selection for Vector Autoregressions." *Review of Economics and Statistics* 97 (2): 436–51.

Giannone, Domenico, Michele Lenza, and Giorgio E. Primiceri. 2019. "Priors for the Long Run." *Journal of the American Statistical Association* 114 (526): 565–80.

Gordon, Robert J. 2007. *The American Business Cycle: Continuity and Change*. University of Chicago Press.

Gross, Daniel P., and Bhaven N. Sampat. 2023. "America, Jump-Started: World War II R&D and the Takeoff of the US Innovation System." *American Economic Review* 113 (12): 3323–56.

Gruber, Jonathan, and Simon Johnson. 2019. *Jump-Starting America: How Breakthrough Science Can Revive Economic Growth and the American Dream*. PublicAffairs.

Guerron-Quintana, Pablo A., and Ryo Jinnay. 2019. "Financial Frictions, Trends, and the Great Recession." *Quantitative Economics* 10 (2): 735–73.

Hall, George J., and Thomas J. Sargent. 2011. "Interest Rate Risk and Other Determinants of Post-WWII US Government Debt/GDP Dynamics." *American Economic Journal: Macroeconomics* 3 (3): 192–214.

Hall, Robert E. 2009. "By How Much Does GDP Rise If the Government Buys More Output?" *Brookings Papers on Economic Activity* 40 (2): 183–249.

Hornbeck, Richard, and Martin Rotemberg. 2021. "Railroads, Market Access, and Aggregate Productivity Growth." Unpublished.

Ikeda, Daisuke, and Takushi Kurozumi. 2019. "Slow Post-financial Crisis Recovery and Monetary Policy." *American Economic Journal: Macroeconomics* 11 (4): 82–112.

Ilzetzki, Ethan. 2024. "Learning by Necessity: Government Demand, Capacity Constraints, and Productivity Growth." *American Economic Review* 114 (8): 2436–71.

Imbs, Jean M. 1999. "Technology, Growth and the Business Cycle." *Journal of Monetary Economics* 44 (1): 65–80.

Janeway, William H. 2012. *Doing Capitalism in the Innovation Economy*. Cambridge University Press.

Jarociński, Marek, and Albert Marcet. 2014. "Contrasting Bayesian and Frequentist Approaches to Autoregressions: The Role of the Initial Condition." Barcelona GSE Working Paper 776.

Jordà, Òscar. 2005. "Estimation and Inference of Impulse Responses by Local Projections." *American Economic Review* 95 (1): 161–82.

Jordà, Òscar, Moritz Schularick, and Alan M. Taylor. 2017. "Macrofinancial History and the New Business Cycle Facts." *NBER Macroeconomics Annual* 31: 213–63.

Jordà, Òscar, Sanjay R. Singh, and Alan M. Taylor. 2020. "The Long-Run Effects of Monetary Policy." NBER Working Paper 26666.

Kantor, Shawn, and Alexander T. Whalley. 2023. "Moonshot: Public RD and Growth." NBER Working Paper 31471.

Kilian, Lutz, and Helmut Lütkepohl. 2017. *Structural Vector Autoregressive Analysis*. Cambridge University Press.

Kurmann, André, and Christopher Otrok. 2013. "News Shocks and the Slope of the Term Structure of Interest Rates." *American Economic Review* 103 (6): 2612–32.

Leff Yaffe, Daniel. 2020. "The Interstate Multiplier." Unpublished.

Li, Dake, Mikkel Plagborg-Møller, and Christian K. Wolf. 2021a. "Local Projections versus VARs: Lessons from Thousands of DGPs." Preprint, arXiv. https://doi.org/10.48550/arXiv.2104.00655.

Li, Dake, Mikkel Plagborg-Møller, and Christian K. Wolf. 2021b. "Local Projections versus VARs: Lessons from Thousands of DGPs." Unpublished.

Mazzucato, Mariana. 2013. *The Enterpreneurial State: Debunking Public versus Private Sector Myths*. Penguin.

Meltzer, Allan H. 2004. *A History of the Federal Reserve, Volume 1: 1913–1951*. University of Chicago Press.

Montiel Olea, José Luis, and Mikkel Plagborg-Møller. 2021. "Local Projection Inference Is Simpler and More Robust Than You Think." *Econometrica* 89 (4): 1789–1823.

Montiel Olea, José Luis, Mikkel Plagborg-Møller, and Eric Qian. 2022. "SVAR Identification from Higher Moments: Has the Simultaneous Causality Problem Been Solved?" *AEA Papers and Proceedings* 112: 481–85.

Moretti, Enrico, Claudia Steinwender, and John Van Reenen. 2025. "The Intellectual Spoils of War? Defense R&D, Productivity and International Spillovers." *Review of Economics and Statistics* 107 (1): 14–27.

Mountford, Andrew, and Harry Uhlig. 2009. "What Are the Effects of Fiscal Policy Shocks?" *Journal of Applied Econometrics* 24 (6): 960–92.

Müller, Ulrich K. 2012. "Measuring Prior Sensitivity and Prior Informativeness in Large Bayesian Models." *Journal of Monetary Economics* 59 (6): 581–97.

Müller, Ulrich K., and Mark W. Watson. 2020. "Low-Frequency Analysis of Economic Time Series." Princeton University Economics Department Working Paper 2020-13.

Nakamura, Emi, and Jón Steinsson. 2014. "Fiscal Stimulus in a Monetary Union: Evidence from US Regions." *American Economic Review* 104 (3): 753–92.

Nakamura, Emi, and Jón Steinsson. 2018. "Identification in Macroeconomics." *Journal of Economic Perspectives* 32 (3): 59–86.

Owyang, Michael T., Valerie A. Ramey, and Sarah Zubairy. 2013. "Are Government Spending Multipliers Greater during Periods of Slack? Evidence from Twentieth-Century Historical Data." *American Economic Review: Papers and Proceedings* 103 (3): 129–34.

Pezzoni, Michele, Reinhilde Veugelers, and Fabiana Visentin. 2022. "How Fast Is This Novel Technology Going to Be a Hit? Antecedents Predicting Follow-On Inventions." *Research Policy* 51 (3): 104454.

Plagborg-Møller, Mikkel, and Christian K. Wolf. 2021. "Local Projections and VARs Estimate the Same Impulse Responses." *Econometrica* 89 (2): 955–80.

Queralto, Albert. 2020. "A Model of Slow Recoveries from Financial Crises." *Journal of Monetary Economics* 114: 1–25.

Ramey, Valerie A. 2011a. "Can Government Purchases Stimulate the Economy?" *Journal of Economic Literature* 49 (3): 673–85.

Ramey, Valerie A. 2011b. "Identifying Government Spending Shocks: It's All in the Timing." *Quarterly Journal of Economics* 126 (1): 1–50.

Ramey, Valerie A. 2019. "Ten Years after the Financial Crisis: What Have We Learned from the Renaissance in Fiscal Research?" *Journal of Economic Perspectives* 33 (2): 89–114.

Ramey, Valerie A. 2020. "The Macroeconomic Consequences of Infrastructure Investment." NBER Working Paper 27625.

Ramey, Valerie A., and Matthew D. Shapiro. 1998. "Costly Capital Reallocation and the Effects of Government Spending." *Carnegie Rochester Conference on Public Policy* 48: 145–94.

Ramey, Valerie A., and Sarah Zubairy. 2018. "Government Spending Multipliers in Good Times and in Bad: Evidence from US Historical Data." *Journal of Political Economy* 126 (2): 850–901.

Romer, Christina D., and David H. Romer. 2010. "The Macroeconomic Effects of Tax Changes: Estimates Based on a New Measure of Fiscal Shocks." *American Economic Review* 100 (3): 763–801.

Rossi, Barbara, and Sarah Zubairy. 2011. "What Is the Importance of Monetary and Fiscal Shocks in Explaining US Macroeconomic Fluctuations?" *Journal of Money, Credit and Banking* 43 (6): 1247–70.

Sargent, Thomas J., and Paolo Surico. 2011. "Two Illustrations of the Quantity Theory of Money: Breakdowns and Revivals." *American Economic Review* 101 (1): 109–28.

Sims, Christopher A. 1993. "A Nine-Variable Probabilistic Macroeconomic Forecasting Model." In *Business Cycles, Indicators, and Forecasting*, edited by James H. Stock and Mark W. Watson, 179–212. University of Chicago Press.

Sims, Christopher A. 2000. "Using a Likelihood Perspective to Sharpen Econometric Discourse: Three Examples." *Journal of Econometrics* 95 (2): 443–62.

Sims, Christopher A., James H. Stock, and Mark W. Watson. 1990. "Inference in Linear Time Series Models with Some Unit Roots." *Econometrica* 58 (1): 113–44.

Sims, Christopher A., and Harry Uhlig. 1991. "Understanding Unit Rooters: A Helicopter Tour." *Econometrica* 59 (6): 1591–99.

Sims, Christopher A., and Tao Zha. 1998. "Bayesian Methods for Dynamic Multivariate Models." *International Economic Review* 39 (4): 949–68.

Uhlig, Harry. 2004. "What Moves GNP?" Paper presented at the Econometric Society 2004 North American Winter Meetings. San Diego, CA, January 3–5.

Uhlig, Harry. 2005. "What Are the Effects of Monetary Policy on Output? Results from an Agnostic Identification Procedure." *Journal of Monetary Economics* 52 (2): 381–419.

Online Appendix to
"The Long-Run Effects of Government Spending"

by Juan Antolin-Diaz (LBS) and Paolo Surico (LBS and CEPR)

Contents

A Estimation algorithm

To estimate the VAR model, we can write it in matrix form as $\mathbf{Y} = \mathbf{X}\mathbf{B}' + \mathbf{U}$. Denoting T the length of the sample, n the number of variables, and p the number of lags in the VAR, $\mathbf{Y} = (\mathbf{y}_1', \ldots, \mathbf{y}_T')'$ is a $T \times n$ matrix, $\mathbf{X} = (\mathbf{x}_1', \ldots, \mathbf{x}_T')'$ is a $T \times K$ matrix, where $K = np + 1$, and $\mathbf{U} = (\mathbf{u}_1', \ldots, \mathbf{u}_T')'$ is a $T \times n$ matrix. The vector of innovations $\mathbf{u}_t$ is assumed to be independently and identically distributed $\mathcal{N}(0, \mathbf{\Sigma})$.

The NIW family of distributions is conjugate for this class of models. If the prior distribution over the parameters is $NIW(\underline{\nu}, \underline{\mathbf{S}}, \underline{\boldsymbol{b}}, \underline{\mathbf{V}})$, then the posterior distribution over the parameters is $NIW(\overline{\nu}, \overline{\mathbf{S}}, \overline{\boldsymbol{b}}, \overline{\mathbf{V}})$, where $\overline{\boldsymbol{b}} = \texttt{vec}\left(\overline{\mathbf{B}}\right)$, $\overline{\mathbf{V}} = \left(\underline{\mathbf{V}}^{-1} + \mathbf{X}'\mathbf{X}\right)^{-1}$, $\overline{\mathbf{B}} = \overline{\mathbf{V}}\left(\underline{\mathbf{V}}^{-1}\underline{\mathbf{B}} + \mathbf{X}'\mathbf{X}\hat{\mathbf{B}}\right)^{-1}$, $\hat{\mathbf{B}} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Y}$, and $\overline{\mathbf{S}} = \hat{\mathbf{S}} + \underline{\mathbf{S}} + \hat{\mathbf{B}}'\mathbf{X}'\mathbf{X}\hat{\mathbf{B}} + \underline{\mathbf{B}}'\underline{\mathbf{V}}^{-1}\underline{\mathbf{B}} - \overline{\mathbf{A}}'\overline{\mathbf{V}}^{-1}\overline{\mathbf{A}}$, $\hat{\mathbf{S}} = \left(\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}}\right)'\left(\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}}\right)$, and $\overline{\nu} = T + \underline{\nu}$. The NIW posterior distributions defined above can be factored into the following conditional and marginal posterior distributions: $\mathcal{N}\left(\overline{\boldsymbol{b}}, \mathbf{\Sigma} \otimes \overline{\mathbf{V}}\right)$ and $p(\mathbf{\Sigma}|\mathbf{y}) \sim \mathcal{IW}\left(\overline{\mathbf{S}}, \overline{\nu}\right)$. This structure allows to independently draw from the posterior distributions.

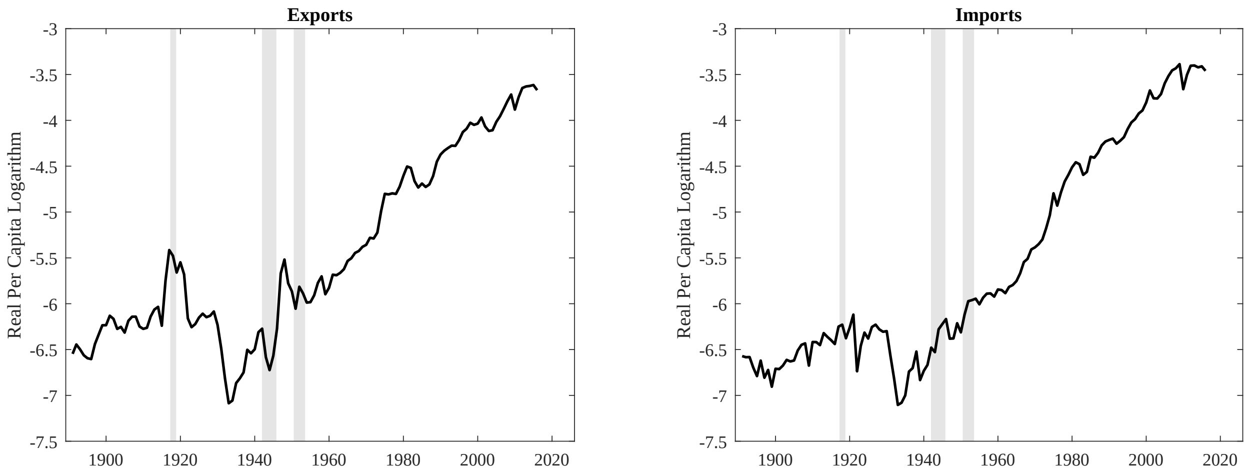
B Data construction

Our starting point is the data set put together by Ramey and Zubairy (2018a) (Ramey and Zubairy, 2018b), which contains seven variables over the sample from 1890Q1 to 2015Q4: the present discounted value of military news (Ramey, 2011), government spending, real GDP, the log GDP deflator, the short-term interest rate, the surplus-to-GDP ratio and the Debt-to-GDP ratio. We use two main transformations of the data. Either we express real GDP per-capita and real government spending per-capita in logarithm or, following Ramey and Zubairy (2018a), we scale them by a measure of GDP trend, estimated as a sixth-degree polynomial for the log of GDP, from 1889q1 through 2015q4, excluding 1930Q1–1946Q4. We extend the dataset in Ramey and Zubairy (2018a) in a number of dimensions that we describe in turn.

Imports and Exports. We use quarterly data from the Bureau of Economic Analysis National Income and Product account on nominal exports and imports of goods and services for the period 1947-2015. For the period 1929-1946, we use annual NIPA data for the same components, exploiting quarterly series on imports and exports from the NBER Macrohistory Data (National Bureau of Economic Research, 1997), to interpolate the annual series using the method in Chow and Lin (1971). For the period 1890-1928, we use the NBER data directly, which we seasonally adjust using the X13 package. The resulting series are displayed in Figure B.1.

Private Investment. Private investment is based on unpublished annual estimates of investment by the Bureau of Economic Analysis, available since 1901 (U.S. Bureau of Economic Analysis, [YYYY]). Before that, we rely on the Macrohistory Database of Jordà et al. (2017) (JST), which provides the investment-to-output ratio, from which levels of investment can be calculated using GDP estimates. We noticed significant differences in the implied investment-output ratios between the BEA and the JST database, so we only use the latter prior to 1901 when no other source is available (Figure B.2, left panel). We then interpolate the annual series to quarterly frequency in the following way: from 1919-1940, we exploit the quarterly series for investment from Gordon (2007) (National Bureau of Economic Research, 1986) to interpolate the annual series using the method in Chow and Lin (1971). For the period when it is available (1889-1918 and 1941-1946), we use

Figure B.1: U.S. EXPORTS AND IMPORTS: 1890-2015



Notes. Exports and Imports are deflated using the GDP deflator. All variables are scaled by the civilian population. Shaded areas represent the three major wars in the sample: World War I, World War II, and the Korean War.

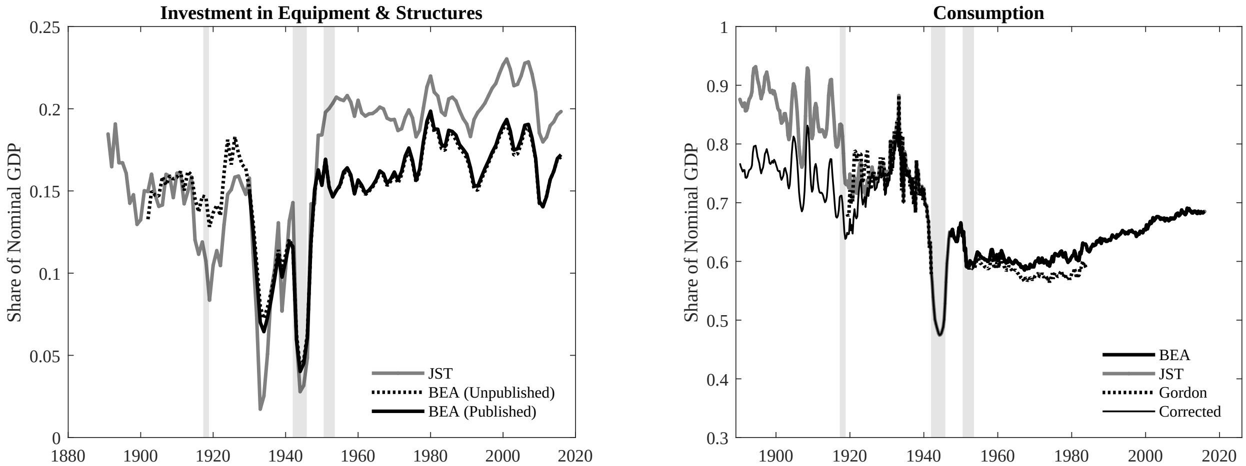
quarterly real GDP from Ramey and Zubairy (2018a) (Ramey and Zubairy, 2018b) to perform the interpolation. The resulting series is visible in the middle panel of Figure B.3.

Private Consumption. We use quarterly data from the Bureau of Economic Analysis National Income and Product account on real consumption of goods (including durables) and services for the period 1947-2015. For the period 1929-1946, we use annual NIPA data for the same components. Before that, we rely on the Macrohistory Database of Jordà et al. (2017). These authors provide an index series of real per capita consumption, which we multiply by population and the CPI to convert back to nominal, and then rebase to the 1929 level of the official NIPA series. We refer to the logarithm of the resulting estimates of consumption as c_t^{JST}. Proceeding in this way, we noticed that the trend growth of consumption for the period 1890-1928 appears understated, resulting in an implausibly large share of consumption to GDP (left panel of Figure B.2). We therefore applied a correction using the national accounts identity:

$$C + \Delta BI = Y - G - I - (X - M) \tag{1}$$

where ΔBI, the change in business investments, drives a wedge between consumption and the right hand side of the equation. Because ΔBI is by definition stationary, however, the right hand side should have the same low frequency trend as consumption itself. Denoting $\bar{c}_t^{JST}$ the Hodrick-

Figure B.2: U.S. CONSUMPTION AND INVESTMENT RATIOS: 1890-2015



Notes. Shaded areas represent the three major wars in the sample: World War I, World War II, and the Korean War.

Prescott trend (smoothing parameter $\lambda = 1000$) of our initial consumption estimate, and $\bar{c}_t^{NA}$ the Hodrick-Prescott trend (smoothing parameter $\lambda = 1000$) of the right hand side of equation (1), our favourite estimate of the consumption time series is:

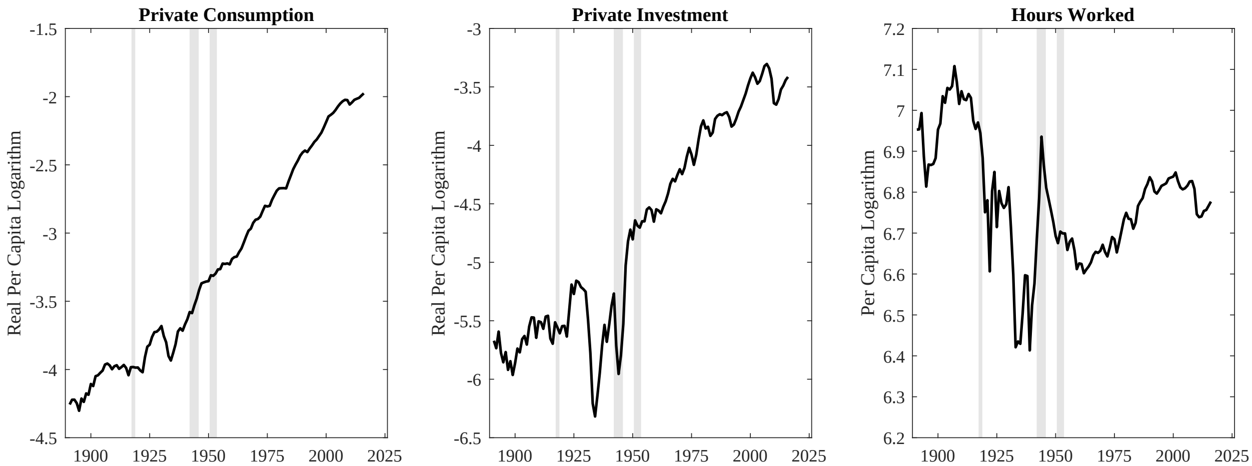
$$c_t^{JST} - \bar{c}_t^{JST} + \bar{c}_t^{NA} \tag{2}$$

The correction (which is only applied for the period 1890-1928) results in more plausible values for the consumption-output ratio and brings it closer to the earlier estimates by Gordon (2007) (National Bureau of Economic Research, 1986), as seen in the right panel of Figure B.2.

Finally, we interpolate the annual series to quarterly frequency in the following way: from 1919-1940, we exploit the quarterly series for investment from Gordon (2007) (National Bureau of Economic Research, 1986) to interpolate the annual series using the method in Chow and Lin (1971). For the period when this are is available (1889-1918 and 1941-1946), we use quarterly real GDP from Ramey and Zubairy (2018a) (Ramey and Zubairy, 2018b) to perform the interpolation. The resulting series is reported in the left panel of Figure B.3.

Nominal interest rates. We first extend backwards the time series for the short-term nominal interest rate, using data from Welch and Goyal (2008a) (Welch and Goyal, 2008b) for the New York Fed commercial paper rate. We obtain the long-term (10-year) interest rate from the same source.

Figure B.3: U.S. CONSUMPTION, INVESTMENT, AND HOURS: 1890-2015

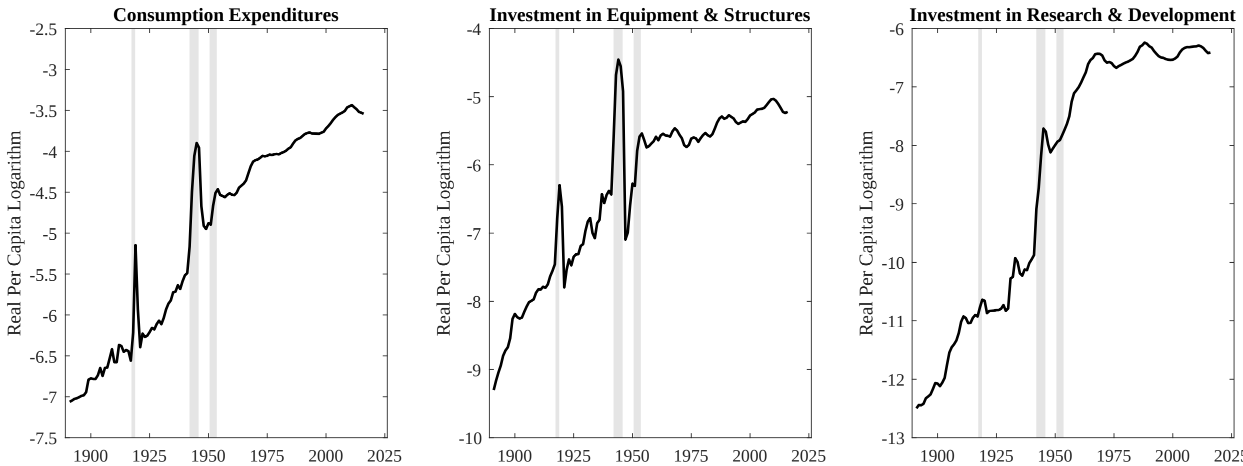


Notes. Private consumption and investment are deflated using the GDP deflator. All variables are scaled by the civilian population. Shaded areas represent the three major wars in the sample: World War I, World War II, and the Korean War.

Public spending components. We also construct new time series that break down government spending into its consumption and investment components. Annual series of government investment are available from the BEA since 1914, but we found that, because of rounding, they are inaccurate until the official NIPA estimates start in 1929. Therefore, we reconstruct the series of public investment and its components for the period 1890-1929 by manually transcribing detailed government outlays data from both the *Historical Statistics of the United States* (Bureau of the Census, 1890-1929) and the annual *Statistical Abstracts* published by the Bureau of the Census (Bureau of the Census, 1949). We transcribe separately data for Federal and State and Local investment. First, the *Historical Statistics*, Chapter P, p.314, provides data points for State and Local "capital outlays" for the years 1890, 1902, 1913. We linearly interpolate observations between these years. For Federal investments in each year between 1890 and 1929, the *Statistical Abstracts* provides detailed annual breakdowns of federal government expenditures by use over the prior ten years. We transcribe this breakdown and sum up all categories of each department that appear to refer to investment, either in Equipment & Structures or in Research & Development.

To classify R&D investments, we rely on the narrative evidence in Bush (1954) and Dupree (1986) to allocate amounts across government departments. In particular, we cross check that total R&D spending matches the estimates reported by (Dupree, 1986, pp. 331-333). We also cross check that the sum of these categories is a good match to the official total amount for the years when

Figure B.4: MAIN CATEGORIES OF U.S. GOVERNMENT SPENDING: 1890-2015



Notes. All components of government spending are deflated using the GDP deflator and scaled by the civilian population. Shaded areas represent the three major wars in the sample: World War I, World War II, and the Korean War.

they overlap. These estimates refer to the year ending on June 30, and thus we average with the next year to obtain an approximation of spending on the calendar year ending in December. After adding the Federal total to the State and Local investment constructed above, we obtain an annual investment series for the total government sector for 1890-1930, which we splice with the official BEA estimate starting in 1914. We then interpolate to quarterly frequency using the series of total government spending, and finally back out government consumption as a residual. Figure B.4 displays the three resulting series for government spending components, in real, per capita terms.

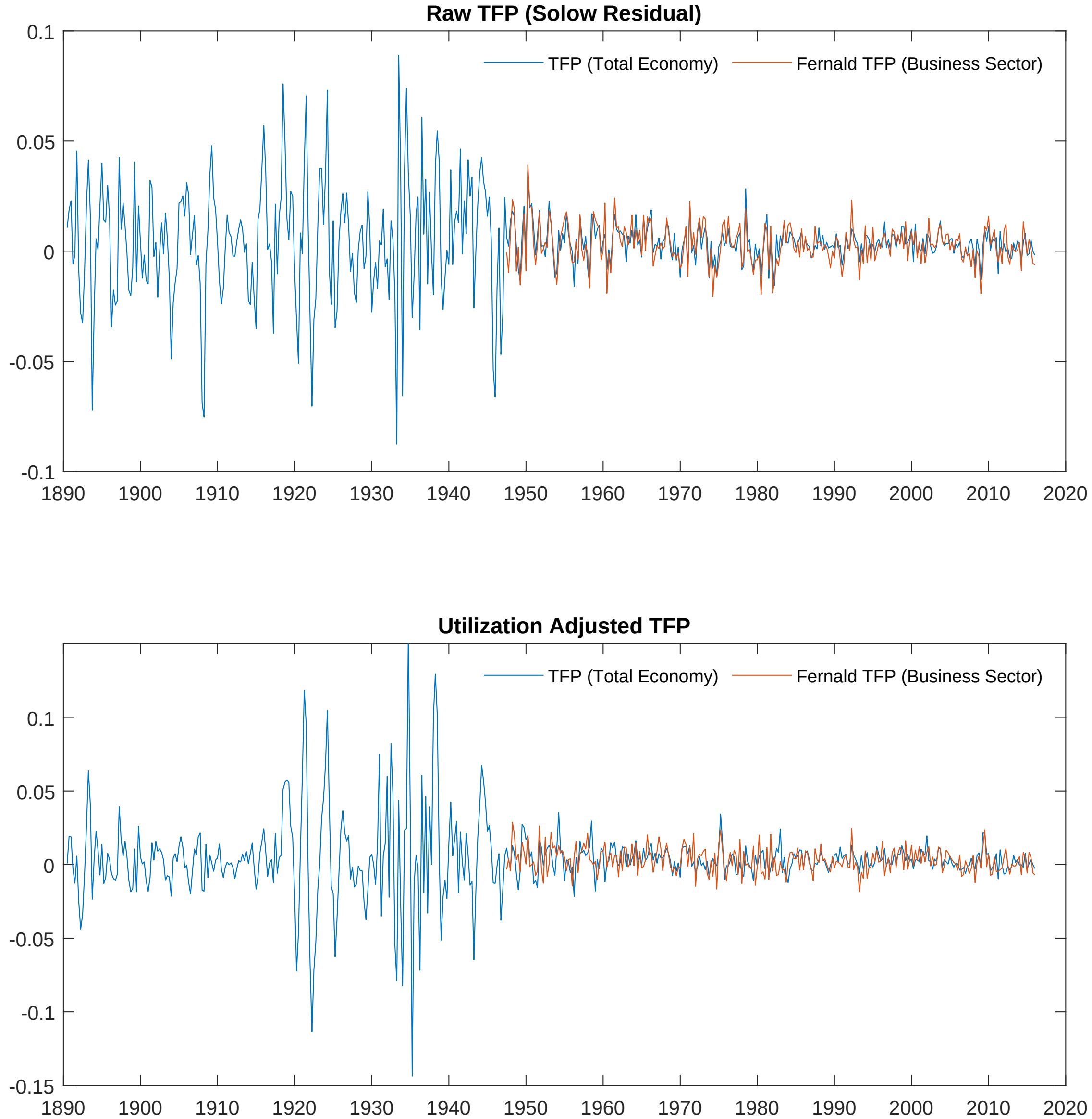
Finally, the quarterly time series for Total Factor Productivity (TFP) has been constructed in two steps. First, we obtain annual measures of hours worked and the capital stock from Bergeaud et al. (2016b) (Bergeaud et al., 2016a).[1] These annual time series are interpolated to quarterly frequency. In the case of investment, we interpolate the annual measure of capital stock using the quarterly series of investment constructed above, cumulated using the perpetual inventory method.[2] For hours, we interpolate the annual measure using the unemployment rate series in Ramey and Zubairy (2018a) (Ramey and Zubairy, 2018b). The raw TFP series is then calculated as the Solow residual using quarterly real GDP, hours worked and the capital stock, assuming a Cobb-Douglass production function with constant returns to scale and a capital share of $\alpha = 0.28$. Second, to derive a measure of TFP adjusted for both capital and labour utilization, we use the method described

[1]We are thankful to Antonin Bergeaud for sharing this data with us.
[2]We assume a depreciation rate of $\delta = 0.1$ per annum.

by Imbs (1999) (and also employed by Jordà et al., 2020). This involves calculating steady-state measures of the capital-labor ratio, the consumption-output ratio and hours. We do so by applying the Hodrick-Prescott filter with a smoothing parameter of $\lambda = 1600$.

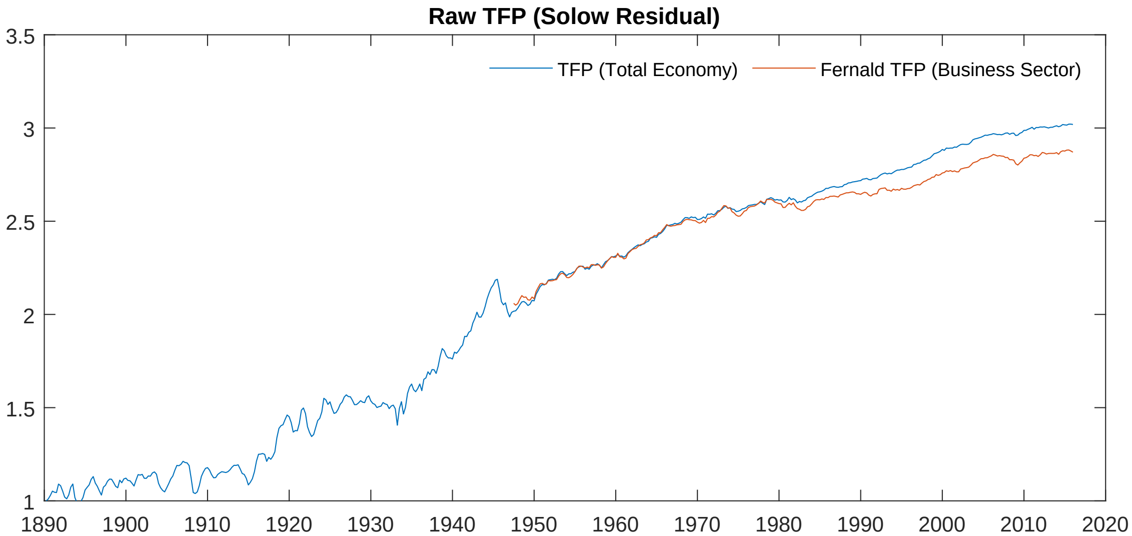
As shown in Figure B.5, which displays growth rates, and Figure B.6, which depicts log-levels, our historical quarterly time series of adjusted TFP, which refers to the whole economy, moves very closely to the more sophisticated and more data intensive measure proposed by Fernald (2012), which covers the business sector only, over the sample in which the two series overlap. Finally, and mostly for completeness, in B.7, we report the quarterly measure of utilization adjusted TFP together with the quarterly time series of military spending news from Ramey and Zubairy (2018a) (Ramey and Zubairy, 2018b). It is interesting to note that our measure of total factor productivity tends to increase persistently after major episodes of military spending buildup, such as the two World Wars and –to a lesser extent– the Korean war, in a way that is visually apparent already at the naked eye. The estimates of our VAR(60) in the main text confirms formally this leading correlation.

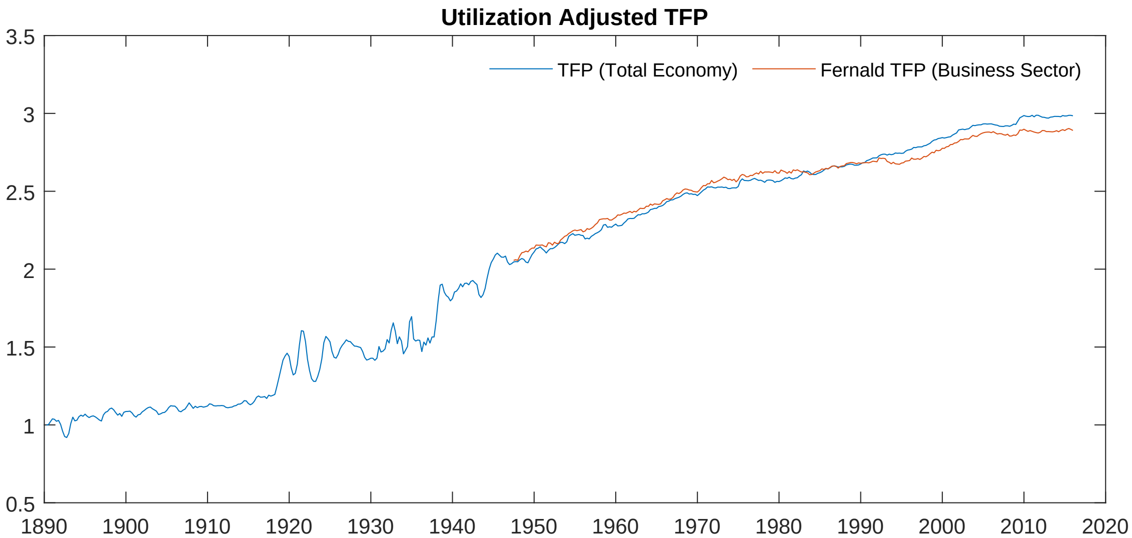
Figure B.5: Raw and Utilization Adjusted TFP Growth Rates



Notes. TFP Growth Rates as described in the Text. Top (bottom) row refers to the raw (utilization adjusted) TFP series.

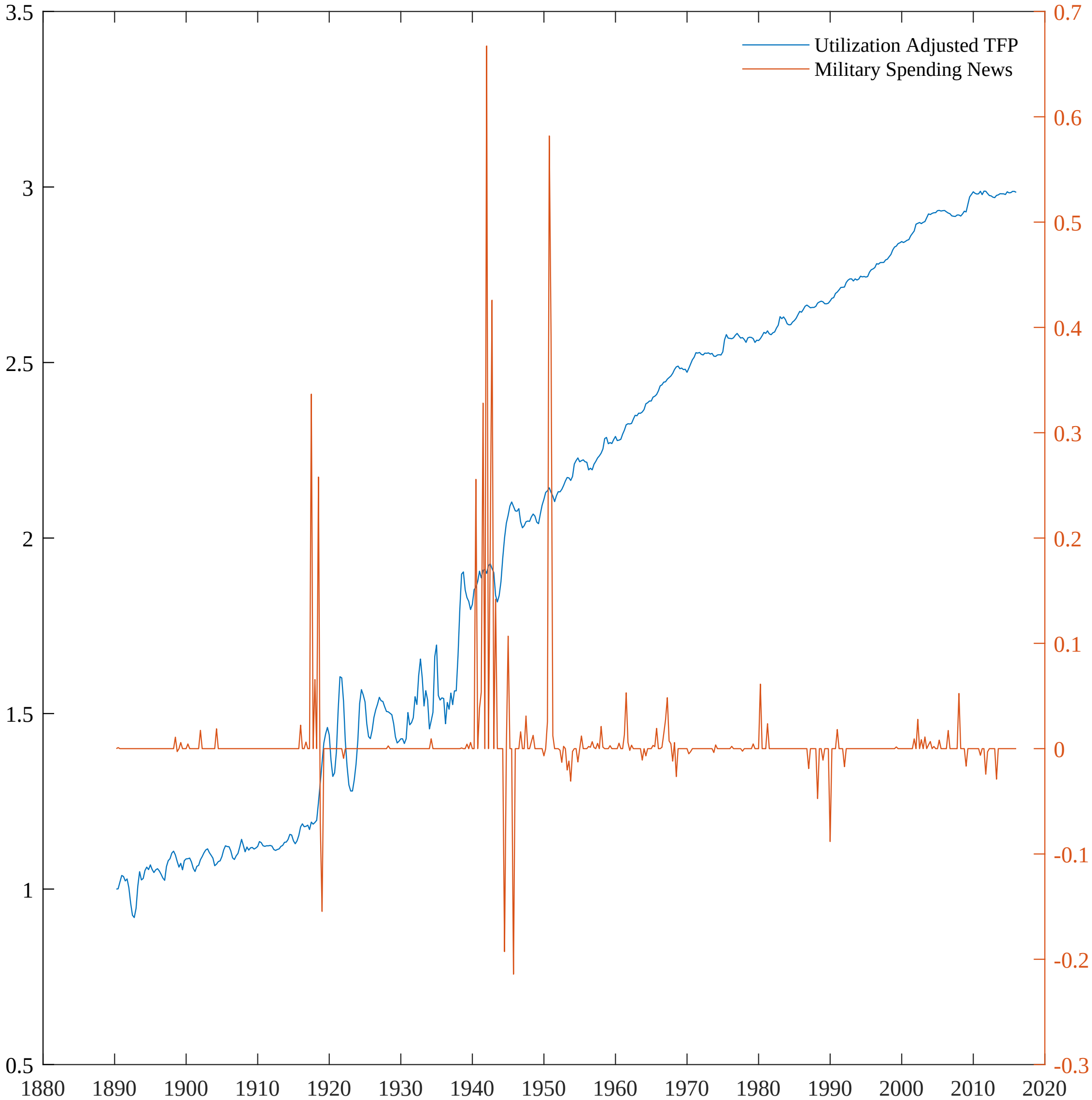
Figure B.6: RAW AND UTILIZATION ADJUSTED TFP LEVELS





Notes. TFP levels as described in the Text. Top (bottom) row refers to the raw (utilization adjusted) TFP series.

Figure B.7: UTILIZATION ADJUSTED TFP LEVELS AND MILITARY SPENDING NEWS

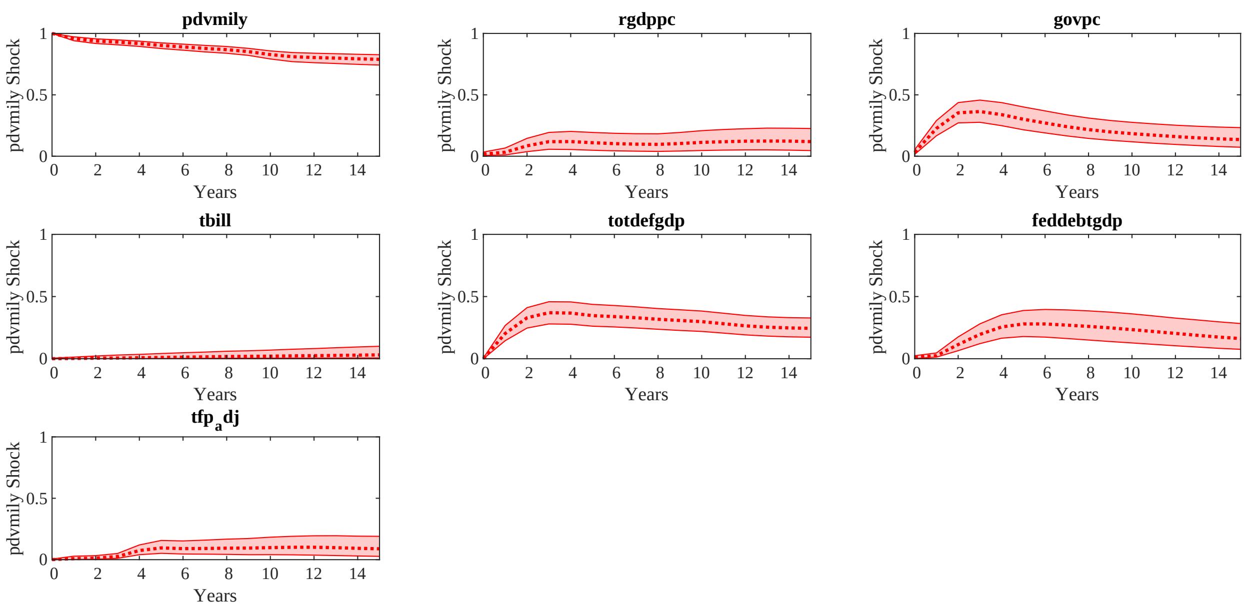


Notes. Utilization-adjusted TFP levels as described in the Text. The military spending news as a percentage of GDP (right axis) is from Ramey and Zubairy (2018a) (Ramey and Zubairy, 2018b)

C Forecast Error Variance Decomposition

In Figure C.1, we report the Forecast Error Variance Decomposition (FEVD) for the baseline results of Figure 1. The darker (lighter) shaded area represents the central 90% posterior credible set.[3] The darker solid line stands for the median estimates. At business-cycle frequencies, the military spending news shock explains about 30%-40% of the variance of the unexpected movements in government spending, whereas it explains about 10% of the variance of real GDP and productivity.

Figure C.1: Forecast Error Variance Decomposition for Military News Shock



Notes. The FEVD is based on an estimated VAR with sixty lags of military spending news, government spending, real per-capita GDP, GDP deflator, short-term interest rate, fiscal deficit to GDP ratio and government debt to GDP ratio. Military spending news is ordered first in the Cholesky factorization. Output, government spending, and the GDP deflator enter the VAR in log-levels. The size of the shock is normalized such as to increase government spending by 1 percent of GDP over the first year after the shock, based on a median G/Y ratio of 19%. The darker (lighter) shaded area represents the 90% HPD interval. The dotted line stands for the median estimates. Results are based on 5000 posterior draws.

[3]The posterior bands of the FEVD in Figure C.1 do not account for possible measurement errors in the instrument.

D Further Results on Sensitivity to prior tightness

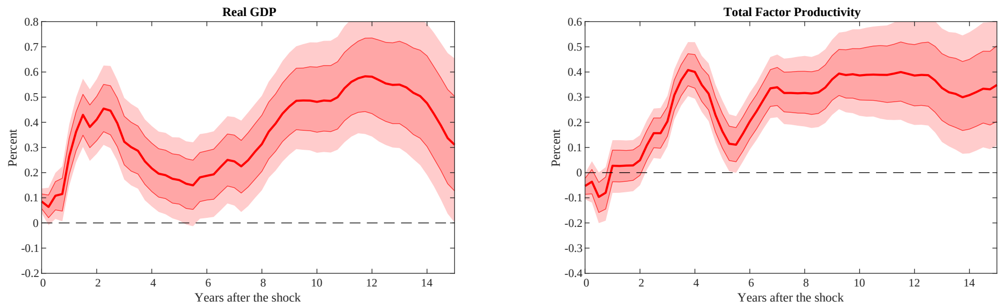
In this section, we present the impulse responses of output and productivity based on variants of the VAR(60) estimated in the main text. The four versions differ in the tightness of the prior hyper-parameters λ (which controls the tightness of the "Minnesota" prior) and θ (which controls the tightness of the "sum of coefficients" prior). The results are reported in Figures D.1 and Figure D.2 below.

In Figure D.1, we perform the analysis that varies the hyper-parameter λ while keeping fixed θ at the baseline value, whereas, in Figure D.2, we conduct the opposite exercise: we vary the hyper-parameter θ while keeping λ fixed at the value estimated using the method in Giannone et al. (2015). Each of the rows starts with relatively uninformative priors, which become progressively tighter going down the figure.

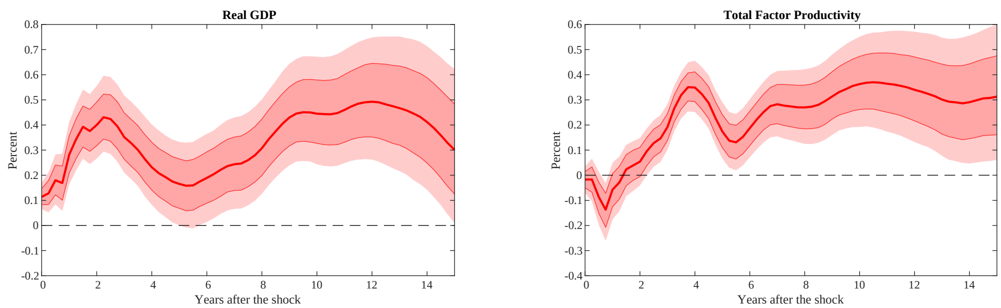
Three main results emerge from this sensitivity analysis. First, progressively tighter priors on the λ hyperparameter of the VAR(60), which are visible going down the rows of Figure D.1, are associated both with smoother shapes of the impulse responses and also with progressively smaller effects at long horizons. Second, despite this progressively increase in tightness, it is still the case that government spending has non-negligible and significantly persistent effects on output and productivity, even in the most conservative specification of $\lambda = 0.1$ in the fourth row. Third, the tightness of the prior hyperparameter θ has some effect on the magnitude of the output and TFP responses at the 15 year horizon in Figure D.2, with stronger effects associated with tighter priors. This is the case, for instance, for the baseline $\theta = 0.01$, which is estimated by maximizing the marginal likelihood as in Giannone et al. (2015). The overall shape and significance of the responses, however, are unchanged even with the relatively loose prior of $\theta = 1$. As already emphasized in the main text, it is very hard for any time series model to draw inference at extremely low frequencies. In a sample of about 500 quarterly observations, like ours, looking at horizons 60 quarters ahead relies on no more than eight non-overlapping samples of 15 years each. Accordingly, we encourage the reader to resist the temptation of drawing inference on whether the effects of government spending may be significant or not after 15 years. Our favourite interpretation of our main finding is instead that the effects of government spending on output and productivity extend significantly beyond business-cycle frequencies, traditionally defined as frequencies beyond eight years.

Figure D.1: Impulse Response Functions Under Alternative Tightness of Prior

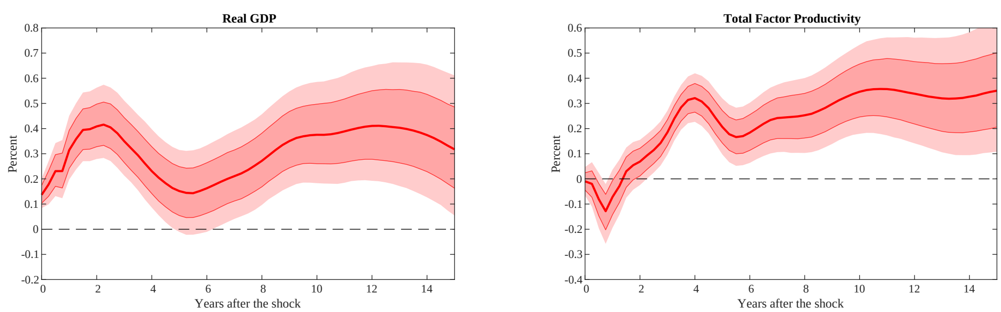
(a) $\lambda = 1$



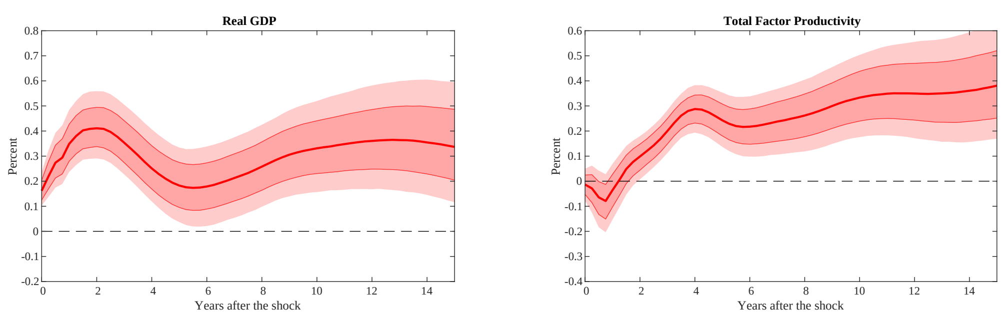
(b) $\lambda = 0.4$



(c) $\lambda = 0.2$



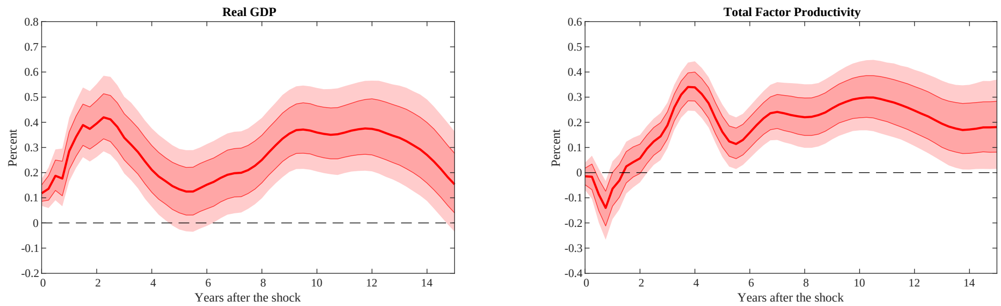
(d) $\lambda = 0.1$



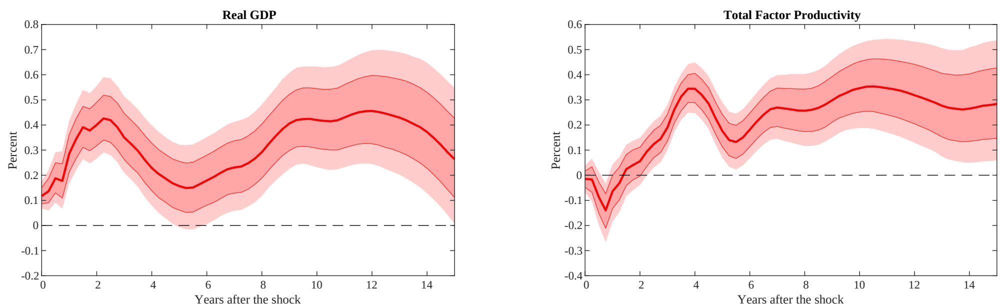
Notes: The solid lines represent the median posterior response. The darker (lighter) shadow area represents the 68^{th} (90^{th}) HPD interval. All specifications use sixty lags of military spending news, government spending, GDP, GDP deflator, treasury bills, deficit to GDP ratio and debt to GDP ratio as in the baseline model. In each row, the parameter λ that governs the tightness of the Minnesota prior in equation (4) takes a different value, ranging from 1 in the top row, to 0.4 and 0.2 in the middle rows, and finally 0.1 in the bottom row. In all cases, the prior hyperparameter θ for the "single unit root" dummy is set at the baseline value of $\theta = 0.001$ that we use as baseline specification in the main text.

Figure D.2: Impulse Response Functions Under Alternative Tightness of Prior

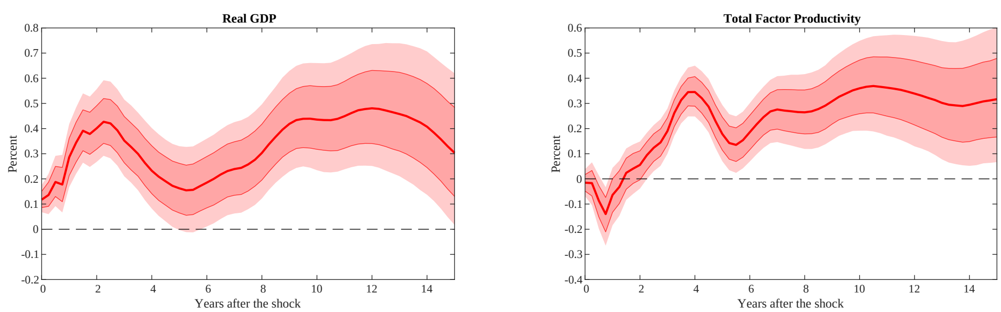
(a) $\theta = 1$



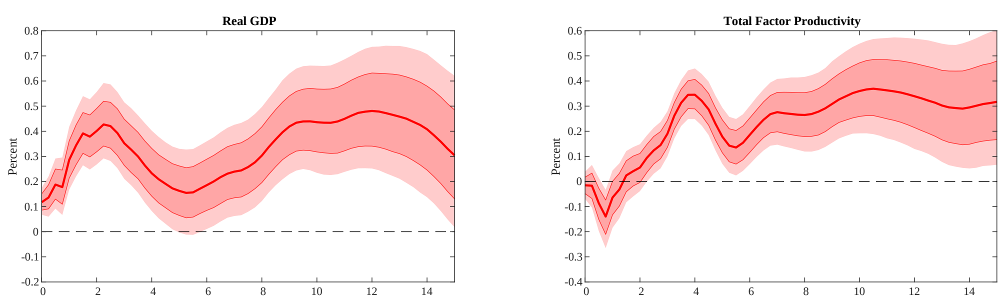
(b) $\theta = 0.1$



(c) $\theta = 0.01$



(d) $\theta = 0.001$



Note: The solid lines represent the median posterior response. The darker (lighter) shadow area represents the 68^{th} (90^{th}) HPD interval. All specifications use sixty lags of military spending news, government spending, GDP, GDP deflator, treasury bills, deficit to GDP ratio and debt to GDP ratio as in the baseline model. In each row, the parameter θ that governs the tightness of the "single unit root" prior in equation (4) takes a different value, ranging from 1 in the top row, to 0.1 and 0.01 in the middle rows, and finally 0.001 in the bottom row. In all cases, the prior hyperparameter λ for the Minnesota prior is set at the baseline value of $\lambda = 0.44$ that we use as baseline specification in the main text.

E Further Details on the Marginal Likelihood and on Estimation with Hierarchical Priors

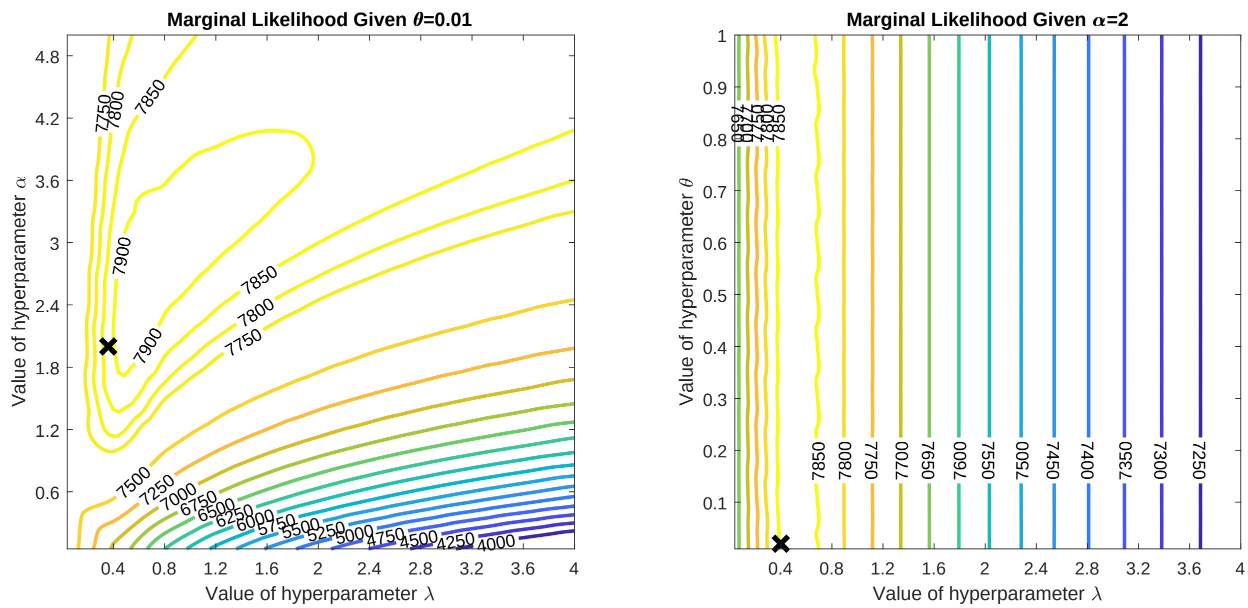
In this section we provide further details on the marginal likelihood of the model and the estimation with hierarchical priors. Our approach follows Giannone et al. (2015) and uses exactly their settings, maximization algorithm, and hyperparameter choices.

In Figure E.1, we report the value of the marginal likelihood as a function of different values of the hyperparameters λ, α and θ. Recall that the joint maximization algorithm leads to $\lambda = 0.36$, $\alpha = 2$, and $\theta = 0.01$. To allow visualization, in the first subplot we fix $\theta = 0.01$, and then vary λ and α, whereas in the second plot we fix $\alpha = 2$, and then vary θ and λ. Therefore, these plots are necessarily a conditional view of the likelihood. The joint maximum is marked with an "X" symbol in both subplots.

A comparison of the two plots indicates that for high values of α (i.e. aggressive discounting of distant lags), relaxing the value of λ does not appear to have much effect on the marginal likelihood. On the other hand, for low values of α, tightening λ leads to sizable improvement of the marginal likelihood. Perhaps unsurprisingly, the two parameters look like substitutes. In the second plot, we report that conditional on $\alpha = 2$, the likelihood is relatively flat as a function of the hyperparameter θ, although a peak is achieved at low values near the optimal value of $\theta = 0.01$.

Estimation with hierarchical priors Finally, we conduct full Bayesian inference on these hyperparameters by specifying a hierarchical prior distribution. We treat the hyperparameters as random variables for which we can elicit prior distributions and conduct posterior inference (Giannone et al., 2015). As discussed in the main text, in this section we also allow for the "sum of coefficients prior", which in principle is inconsistent with cointegration and thus was excluded from the baseline specification for theoretical reasons. Therefore, we have an extra hyperparameter, μ, which governs the tightness of this prior. We follow Giannone et al. (2015) in choosing as hyperpriors for λ, μ, θ and α Gamma densities with modes of 0.2, 1, 1 and 2 —the values recommended by Sims and Zha (1998)— and standard deviations of 0.4, 1, 1 and 1, respectively.

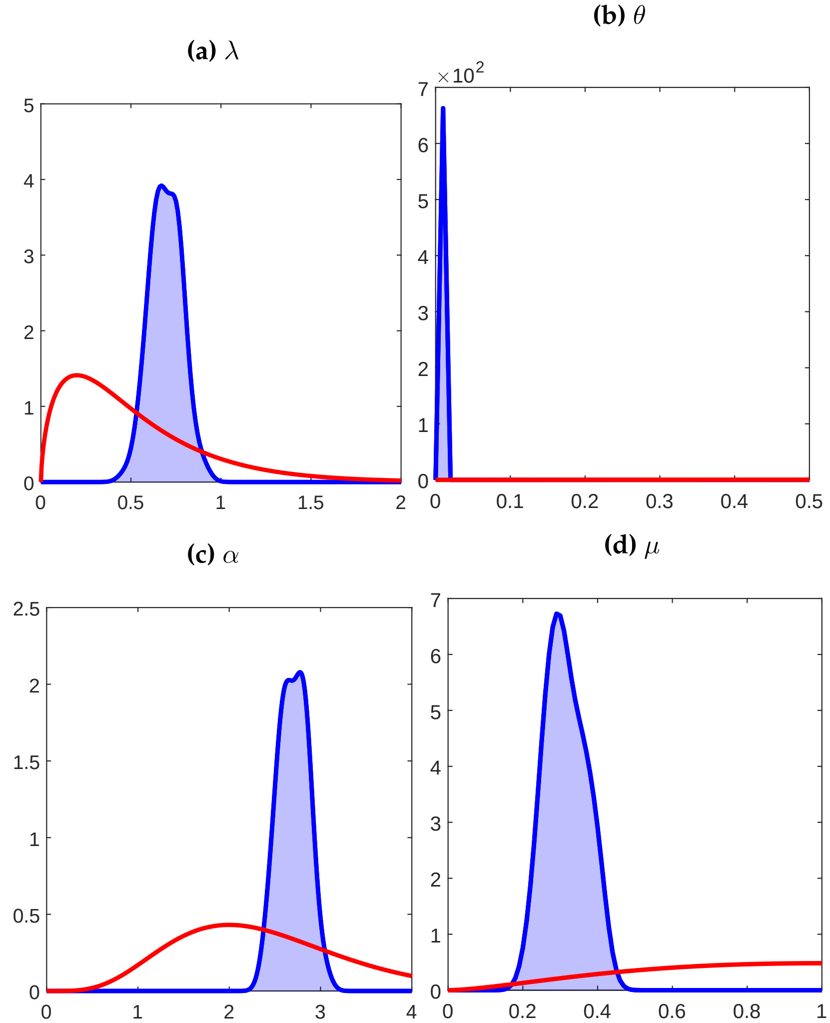
Figure E.1: MARGINAL LIKELIHOOD AS A FUNCTION OF PRIOR HYPERPARAMETERS



Notes: This chart report the contours of the marginal likelihood as function of two hyperparameters at each time. In the left panel, we fix $\theta = 0.01$, and vary α on the vertical axis and λ on the horizontal axis. In the right panel, we set $\alpha = 2$, and vary θ on the vertical axis and λ on the horizontal axis. The symbol "X" represents the joint maximum in each plot.

In Figure E.2, we report the prior distributions (in red) and posterior distributions (in shaded blued) of the hyperparamaters of the hierarchical priors specification that we employ for the impulse response function analysis of Figure 5. As can be seen, the priors are relatively flat and provide a minimum amount of information. Nevertheless, the posterior distribution appears well behaved. In line with the results reported above, we obtain values of the hyperparameters that are consistent with a moderately tight degree of shrinkage, except for the "single unit root" prior, for which a tighter value of θ is preferred.

Figure E.2: Prior and Posterior Distribution of Hyperparameters



Notes: This chart reports the prior distributions (in red) and posterior distributions (in shaded blued) of the hyperparamaters of the hierarchical priors specification behind the results in Figure 5. See Notes to Figure 5 for further details.

F Not Always Long-Run Effects

A possible concern is that the rich parameterization may have introduced some spurious cycles in the VAR(60) reduced-form estimates. Alternatively, a propagation mechanism a la Comin and Gertler (2006) or Beaudry et al. (2020) may drive such a large share of low-frequency variation in the data that any shock would produce highly persistent dynamics. In either case, it would be misleading to infer that government spending is responsible for the estimated long-lasting effects on output and productivity.

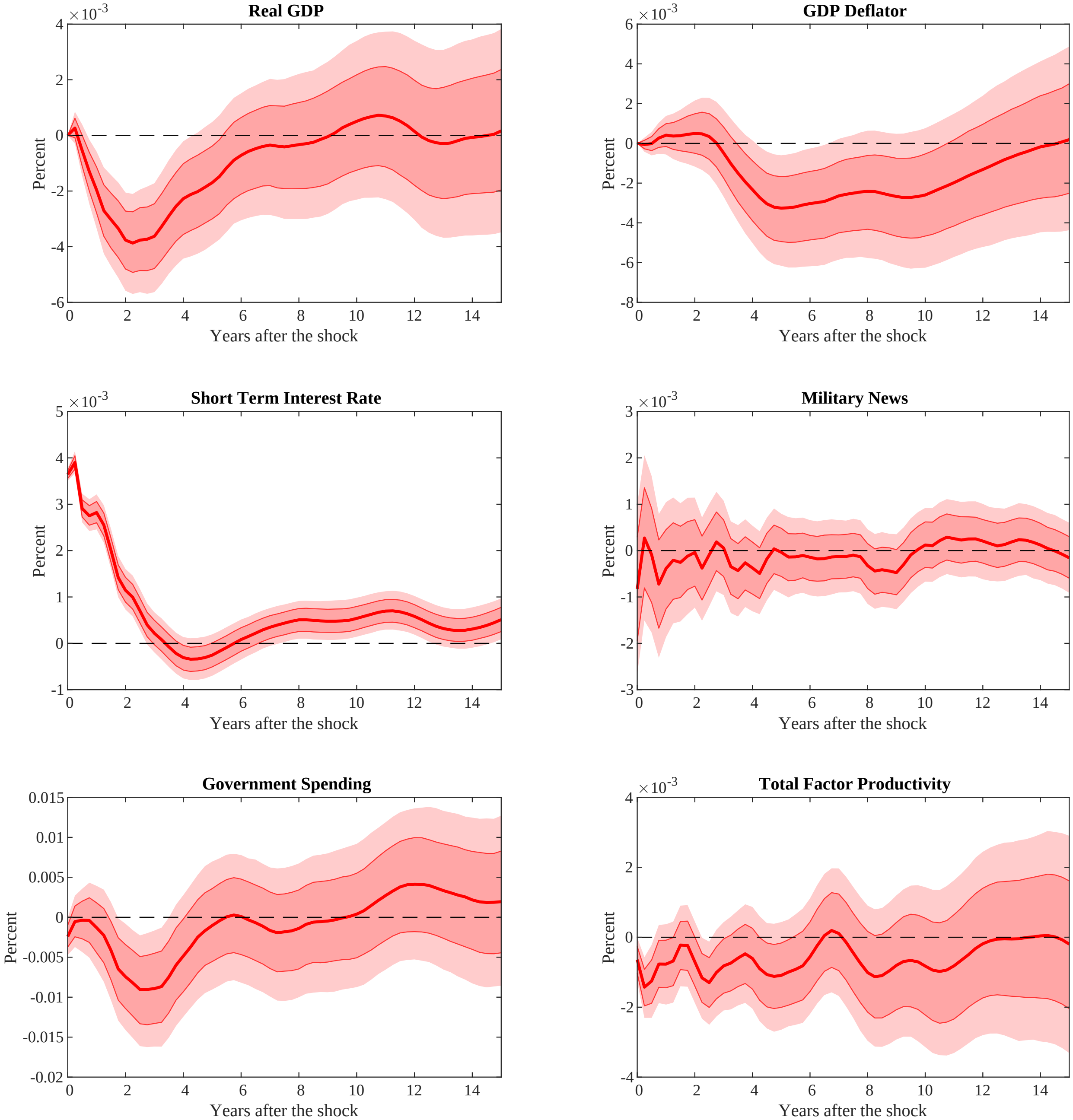
To evaluate this hypothesis, we augment our baseline VAR(60) with the GDP deflator and identify the monetary policy shock via a Cholesky factorization in which real GDP and the GDP deflator are ordered before the short-term interest rate. The idea behind this identification, which has a long tradition in empirical macro (Christiano et al., 2005), is that while monetary policy responds contemporaneously to changes in output and prices, it takes at least a quarter for the effects of central bank interventions to transmit to the economy.[4] It is worth emphasizing that the purpose of this exercise is to verify whether a different orthogonal shock would also produce persistent movements in output and productivity at long horizons. As such, the specific restrictions that are imposed to identify such a shock (and thus its economic interpretation) are not really crucial for our purposes.

The estimated impulse responses to a monetary policy shock are presented in Figure F.1 and they closely resemble those typically found in the empirical monetary literature (Christiano et al., 2005). The estimates of this structural VAR(60) point to significant short-term contractions in output and productivity but exhibit no second wave of effects at longer horizons.[5] We conclude that the highly persistent effects that we have documented in this paper are likely to reflect a genuine (low-frequency) feature of the U.S. government spending data rather than an artifact of our richly parameterized model, or a systematic response of output to any type of shocks.

[4]Relative to equally popular approaches such as those based on narrative evidence and the Greenbook forecasts (Romer and Romer, 2004) or on high frequency movements of interest rate futures around policy announcements (Gürkaynak et al., 2005), the recursive identification has the notable advantage of being readily implementable in our long sample, over which neither the Fed internal forecasts nor the interest rate futures are available.

[5]The results in this section are not necessarily inconsistent with those in Jordà et al. (2020). First, these authors look at an international panel of 17 advanced economies whereas we focus on the U.S. only. Second, and most importantly, Jordà et al. (2020) isolate the exogenous component of monetary policy via the trilemma in international finance while we use a more conventional Cholesky identification, whose only purpose is to show one example in which the type of contemporaneous zero restrictions used in the main analysis can produce small and insignificant effects at long horizons.

Figure F.1: Impulse Responses to Monetary Policy Shock



Notes. The impulse responses are based on an estimated VAR with sixty lags of military spending news, government spending, real per-capita GDP, utilization-adjusted TFP, short-term interest rate, fiscal deficit to GDP ratio and government debt to GDP ratio. Government spending, GDP and the GDP deflator enter the VAR in log-levels. Military spending news is ordered first in the Cholesky factorization. The size of the shock is normalized such as to increase government spending by 1 percent of GDP over the first year after the shock, based on a median G/Y ratio of 19%. The darker (lighter) shaded areas represent the central 68% (90%) high posterior density (HPD) intervals. The darker solid lines are the median estimates. Results are based on 5000 posterior draws.

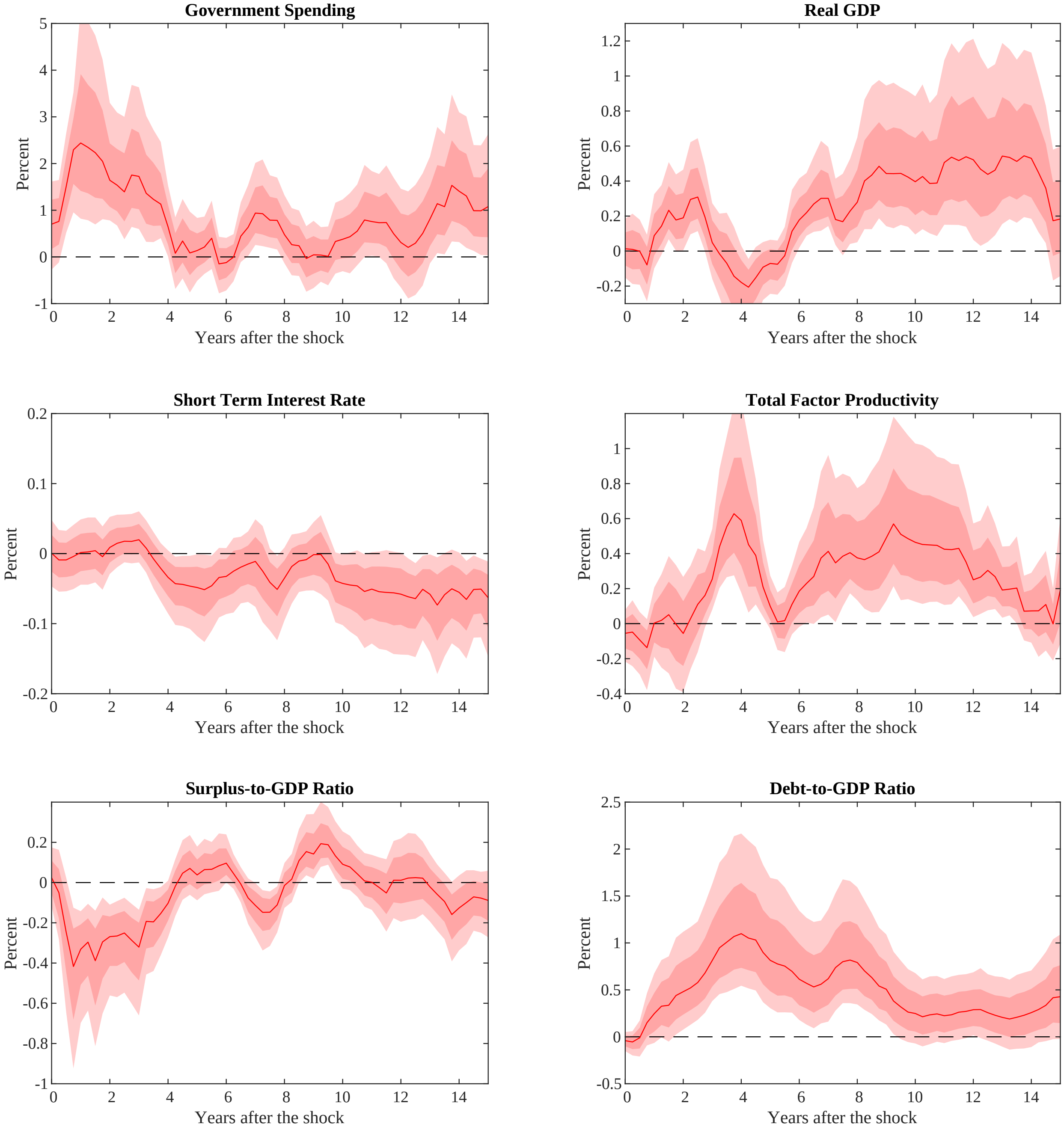
G Local Projections with Principal Components

In this section, we present the results of an alternative strategy to reduce the curse of dimensionality in the context of frequentist inference. This involves collapsing the $60 \times 7 = 420$ original controls into $k = 43$ Principal Components (PCs) that explain the bulk of their variance. It is interesting to note that the first few PCs capture the (common) low frequency components of the 420 variables while the successive PC are more likely to summarize higher frequency covariation. An important question is therefore how to select the number k of principal components to use in the regression. We follow the criteria of Bai (2004), which is robust to unit roots in the data, as it is in our case. We find almost identical number of components whether we use the IPC_1 criterion ($k = 43$) or the IPC_2 criterion ($k = 40$).

The results from this exercise are displayed in Figure G.1 and reveal a number of patterns that are very robust also to using local projections (LPs) with PCs. First, the initial increase in government spending is large and significant before returning to zero after four years. At longer horizons, there is some evidence of another fiscal expansion but the magnitude seems smaller and is less precisely estimated than in the short-run. Second, consistent with the BVAR(60) estimates of Figure 1 and the LP(20) estimates of Figure 6, the response of GDP is characterized by two humps. The first hump is shorter-lived (between years 2 and 4) whereas the second hump lasts longer (between years 6 and 14) and displays a higher peak. Third, productivity displays a delayed significant response that peaks around year 4, before returning to pre-shock levels. After 24 quarters, however, productivity increases again, with effects that appear much more persistent than around the first peak. Fourth, the significant fiscal deficit recorded over the four years after the military spending shock accumulates into a significant increase in the debt-to-GDP ratio, with both patterns reversed at longer horizons.

We conclude that our main finding of large and significant effects of government spending on output and productivity beyond business cycle frequencies (i.e. beyond eight years) seems a very robust feature of U.S. data, which emerges also using local projections with principal components. At horizons around or beyond 15 years, there is no conclusive evidence on whether the long-lasting effects reported throughout the paper are best interpreted as permanent or highly persistent. But this should not come as a surprise. A sample of about 500 quarterly observations, like ours, can

Figure G.1: IRFs to Military News Shock from LP Using Principal Components



Notes. The impulse responses are based on an estimated VAR with sixty lags of military spending news, government spending, real per-capita GDP, utilization-adjusted TFP, short-term interest rate, fiscal deficit to GDP ratio and government debt to GDP ratio. Government spending, GDP and the GDP deflator enter the VAR in log-levels. Military spending news is ordered first in the Cholesky factorization. The size of the shock is normalized such as to increase government spending by 1 percent of GDP over the first year after the shock, based on a median G/Y ratio of 19%. The darker (lighter) shaded areas represent the central 68% (90%) high posterior density (HPD) intervals. The darker solid lines are the median estimates. Results are based on 5000 posterior draws.

only accommodate up to 8 non-overlapping sample of 15 years. It follows that time-series models are unlikely to have enough statistical power to draw accurate inference at very very long horizons (i.e. at or beyond 15 years) based on a relatively short sample.

H The Low-Frequency Covariability of Public R&D and GDP

In this appendix, we shed light on the reduced-form covariability of public R&D and GDP at low frequencies, using the methods proposed by Müller and Watson (2020). We then compare unconditional and conditional forecasts of GDP and government R&D at 25 and 50 year horizons produced by the Low Frequency Factor (LFF) model of Müller and Watson (2020) and a Bayesian VAR(60).

Following Müller and Watson (2020), we consider the following factor model for $\mathbf{x}_t = \{\Delta R\&D_t, \Delta Y_t, \Delta TFP_t\}$:

$$\mathbf{x}_t = \boldsymbol{\mu} + \boldsymbol{\lambda}\mathbf{f}_t + \mathbf{e}_t \tag{3}$$

where $\boldsymbol{\mu}$ are the unconditional means of the series, $\mathbf{f}_t$ denotes the unobserved common factors, $\boldsymbol{\lambda}$ denotes the factor loadings, and $\mathbf{e}_t$ denotes a vector of mutually independent errors that capture the residual variability in the series. Using the notation of Müller and Watson (2020), this can be re-expressed at low frequencies as:

$$\mathbf{X}_T^0 = \boldsymbol{\iota}_{q+1}\boldsymbol{\mu}' + \mathbf{F}_T\boldsymbol{\lambda}' + \mathbf{E}_T \tag{4}$$

where $\mathbf{F}_T = T^{-1}\boldsymbol{\Psi}_T^{0\prime}\mathbf{f}_{1:T}$ and similarly for $\mathbf{E}_T$, with $\boldsymbol{\Psi}_T^{0\prime}$ representing cosine functions with periods 2 through $2/q$. We refer the interested reader to Müller and Watson (2020) for further details. We focus on frequencies lower than 15 years, which correspond to both the lag length of the BVAR and the IRF maximum forecast horizon. In our case, $\mathbf{f}_t$ is a scalar f_t, and therefore the factor loadings λ_i are also scalars.

It is easy to see that in this model, the low frequency covariance between variables i and j is representd by $\text{Cov}(x_i, x_j) = \lambda_i \text{Var}(f_t)\lambda_j$. Following the usual normalization in factor models, we fix the first loading (i.e. the one that corresponds to R&D) to 1. This means that the significance of the low frequency covariability between R&D and GDP, on the one hand, and between R&D and

TFP, on the other hand, can be assessed by simply looking at the factor loadings of GDP and TFP, respectively. In Table H.1, we report the estimation results. These reveal that the factor loadings are both positive and significant, even at the 5^{th} percentile of the posterior distribution, thereby indicating a positive association between R&D, GDP and TFP at frequencies lower than 15 years.

Table H.1: LOW FREQUENCY FACTOR MODEL ESTIMATES

	Mean	5^{th}	16^{th}	50^{th}	84^{th}	95^{th}
Unconditional Means						
$\mu_{R\&D}$	2.69	1.10	1.74	2.70	3.64	4.27
μ_{GDP}	2.06	1.22	1.59	2.05	2.54	2.98
μ_{TFP}	1.71	1.06	1.34	1.69	2.07	2.41
Factor Loadings						
$\lambda_{R\&D}$	1.00	1.00	1.00	1.00	1.00	1.00
λ_{GDP}	2.87	0.79	1.46	2.80	4.30	5.44
λ_{TFP}	1.90	0.46	0.87	1.80	2.97	3.86

Notes: Posterior Estimates of selected parameters of the Low Frequency Factor Model as described in Müller and Watson (2020).

Following Müller and Watson (2020), in each panel of Tables H.2 and H.3, we present results for two type of forecasts. In the first exercise, we compute the distributions of the unconditional forecasts of public R&D and GDP implied by a BVAR(60) in government R&D, GDP and TFP, and the Low Frequency Factor Model described above. The unconditional forecasts are defined $\mathbb{E}[\mathbf{x}_{T+1:T+h}|\mathbf{x}_{1:T}] = \frac{1}{h}\sum_{i=1}^{h}\mathbb{E}\left[\mathbf{x}_{T+i}|\mathbf{x}_1,\ldots,\mathbf{x}_T\right]$ for each model. The BVAR conditional forecasts based on the average growth rate can be readily implemented using the methodlogy described in Antolin-Diaz et al. (2021). These are displayed at the top of Panel A and Panel B of each Table, respectively, with columns representing mean value and different portions of the forecast distributions from the 5^{th} to the 95^{th} percentiles. The second exercise (at the bottom of each panel) refers to the forecasts for GDP *conditional* to an increase in government R&D as large as during the Manhattan project. The reason for this choice is twofold. First, we want to focus on a large public program, as these are more likely to trigger the type of mechanisms and persistent effects that we have highlighted throughout the paper. Second, this magnitude roughly corresponds to a two standard deviation shock of the unconditional forecast of government R&D growth implied by the estimates of the BVAR(60), a metric frequently used in empirical macroeconomic studies.

Table H.2: BVAR (60) VS WATSON AND MULLER (2020)'S LFF MODEL: 25 YEAR HORIZON

Panel A: Bayesian VAR(60)

Horizon (h)	Mean	5^{th}	16^{th}	50^{th}	84^{th}	95^{th}
Unconditional Forecast						
R&D	5.75	-2.85	0.72	5.73	10.95	14.98
GDP	0.95	-0.77	-0.06	0.99	2.00	2.60
Conditional Forecast: $\overline{R\&D}_{T+1:T+100=12.68\%}$						
R&D	12.68	12.68	12.68	12.68	12.68	12.68
GDP	2.19	1.25	1.60	2.18	2.84	3.24

Panel B: LFF model of Watson and Mueller (2020)

Horizon (h)	Mean	5^{th}	16^{th}	50^{th}	84^{th}	95^{th}
Unconditional Forecast						
R&D	2.39	-5.23	-1.85	2.48	6.59	9.69
GDP	1.92	-0.41	0.65	1.94	3.21	4.22
Conditional Forecast: $\overline{R\&D}_{T+1:T+100=12.68\%}$						
R&D	12.68	12.68	12.68	12.68	12.68	12.68
GDP	2.28	0.01	0.97	2.25	3.62	4.63

Notes: For each Panel, the first two rows present unconditional forecasts of GDP and R&D based, respectively on a Bayesian VAR and the Low Frequency Factor Model of Müller and Watson (2020). The second row presents the results of conditioning the average growth rate of R&D over the next 25 years (100 quarters) to equal 12.68%, a magnitude equal to the increase during World War II. BVAR conditional forecasts are implemented following Antolin-Diaz et al. (2021).

There are three main takeaways from Tables H.2 and H.3. First, the unconditional forecasts of government R&D and GDP implied by the BVAR(60) and the method developed by Müller and Watson (2020) are very similar at all percentiles, over both the 25 year horizon (Table H.2) and the

50 year horizon (Table H.3). Second, the distribution of the GDP conditional forecasts based on the LFF model of Müller and Watson (2020) tend to be wider than the conditional distributions implied by the BVAR(60), though the latter is fully (mostly) contained in the upper portion of the former at the 25 (50) year horizon. Third, and more importantly, despite the methods being very different, the BVAR(60) and the factor model proposed by Müller and Watson (2020) share the finding of a significant low-frequency covariability between government R&D and GDP, at either horizons, thereby corroborating the view that our key result is a genuine feature of U.S. data.

Table H.3: BVAR (60) VS WATSON AND MULLER (2020)'S LFF MODEL: 50 YEAR HORIZON

Panel A: Bayesian VAR(60)

Horizon (h)	Mean	5^{th}	16^{th}	50^{th}	84^{th}	95^{th}
Unconditional Forecast						
R&D	7.13	−6.00	−0.73	7.64	14.61	18.80
GDP	1.02	−1.96	−0.68	1.14	2.70	3.57
Conditional Forecast: $\overline{R\&D}_{T+1:T+200=12.68\%}$						
R&D	12.68	12.68	12.68	12.68	12.68	12.68
GDP	2.30	1.32	1.61	2.27	2.96	3.39

Panel B: LFF model of Mueller and Watson (2020)

Horizon (h)	Mean	5^{th}	16^{th}	50^{th}	84^{th}	95^{th}
Unconditional Forecast						
R&D	2.61	−3.82	−0.84	2.72	6.05	8.73
GDP	1.94	0.03	0.93	1.96	2.96	3.79
Conditional Forecast: $\overline{R\&D}_{T+1:T+200=12.68\%}$						
R&D	12.68	12.68	12.68	12.68	12.68	12.68
GDP	2.30	0.44	1.26	2.23	3.36	4.31

Notes: For each Panel, the first two rows present unconditional forecasts of GDP and R&D based, respectively on a Bayesian VAR and the Low Frequency Factor Model of Müller and Watson (2020). The second row presents the results of conditioning the average growth rate of R&D over the next 50 years (200 quarters) to equal 12.68%, a magnitude equal to the increase during World War II. BVAR conditional forecasts are implemented following Antolin-Diaz et al. (2021).

I A Brief Narrative Account of Major Federal R&D Programs

In this Appendix, we provide a brief narrative account of the major public R&D programs funded in the United States over our long historical sample. Although the data includes spending at both the federal and the state and local levels, the discussion below focuses on federal funding towards R&D because it represents about 90% of the total public expenditure on R&D and it underwent major shifts during the XXth century. In contrast, state and local R&D public funds have grown steadily over time and have not experienced abrupt variations.

From the end of the XIX^{th} century to World War I. Dupree (1986) surveys the history of federal investment in Research & Development, from the creation of the United States until the outbreak of World War II. From 1890 to 1940, R&D expenditure represented 1% or less of the total federal budget. Agricultural and natural-resource oriented research, such as the Geological survey and the weather bureau, were far more dominant targets of public spending at the beginning of the XX^{th} century. Indeed, our reconstructed estimates indicate that in 1900, the Department of Agriculture was responsible for 70% of all federal R&D outlays. Its activities included the establishment of weather stations and laboratories, with the objective of preventing disease and improving farm productivity.

The beginning of the XX^{th} century saw the creation of various federal agencies, whose objective was to provide support to business activities and to address national objectives. Examples include the Public Health Service and, within it, the Hygienic Laboratory, predecessor of the National Institutes of Health, established in 1901. In the same year, the National Bureau of Standards (predecessor of the National Institute of Standards and Technology) was established to maintain standards of weights and measures in the face of rapid technological expansion.

World War I and the interwar period. World War I spurred new research efforts, and for the first time defense and national security started rivaling agricultural research. This includes the creation of the National Advisory Committee for Aeronautics, the predecessor of NASA, formed in 1915. There was not, however, a governmental agency for federal R&D with an organization structure similar to the department of Agriculture, with much of the research done in support of the war efforts being coordinated by the National Research Council, an advisory arm of the National Academy of Sciences. In the meantime, social sciences became more prominent, with the Bureau of the Census and the Bureau of Labor Statistics playing an important role within the departments of commerce and labor. During the New Deal era, federal research in health expanded and federal funding to the Public Health Service increased as part of the Social Security Act. A major achievement was the growth of the National Institutes of Health (NIH), established in 1931, and expanded in 1937 with the creation of the National Cancer Institute.

World War II and the Manhattan project. The war constituted a revolution in both the scale and the scope of federal R&D. Just before the United States entered the war, President Roosevelt set up the Office of Scientific Research and Development (OSRD), which was responsible for coordinating R&D efforts in support of the war. Large numbers of academic researchers were mobilized to work in their own institutions' laboratories on wartime R&D projects. This was a key difference with World War I, when scientists working on military projects had been recruited by military agencies. Another important innovation was the establishment of R&D contracts as a mechanism to pay for private performance of work whose approach and outcome could not be specified precisely in advance. Importantly, the federal government agreed to compensate university and industry performers for the indirect or over-head costs of R&D undertaken as part of grants and contracts, in addition to paying for direct expenses. Moreover, to carry out the vastly increased scale of R&D during World War II, major investments were made in government research laboratories (National Research Council, 1995). The largest and most notable of all projects was the Manhattan Engineering District, which was responsible for the development of the atomic bomb. At its peak in 1944, the Manhattan project accounted for nearly one-tenth of all public and private R&D performed in the United States.

In the same year, President Roosevelt asked Vannevar Bush, then director of OSRD, to 'export' the wartime R&D experience to a peacetime institution. The celebrated Bush (1954) report was delivered to President Truman in July 1945. It argued that knowledge and scientific research was an essential ingredient for improving the nation well being, health, economic growth, and national security. Moreover, the report stated that the the federal government had an important responsibility to support both scientific research and the training of new scientific talents. The key recommendation of the report was the establishment of a central research funding agency, initially called the National Research Foundation, to implement those responsibilities.

The Post-WWII Scientific Establishment. After the war, and heavily influenced by the vision laid out by the Bush report, the wartime scientific efforts were consolidated through the creation, after much congressional debate, of the National Science Foundation in 1950. Major increases in R&D efforts, including the creation of DARPA and NASA, followed the Soviet launch of the Sputnik satellite in 1958. This event revealed that the United States had fallen behind the Soviets in space technology. In 1961, president Kennedy kick-started the Apollo program by which NASA landed on the moon in 1969. The conclusion of the Apollo program led to a decline in federal R&D spending, which did not reach its 1960s peak in real per capita terms until the 1980s.

Reagan and the Strategic Defense Initiative. The 1980s witnessed large increases in defense R&D by the Reagan administration, including the Strategic Defense Initiative (SDI, popularly known as 'Star Wars'). This was also motivated by concerns about the Soviet Union and a desire to achieve technological s it uperiority. Defense R&D spending peaked again in 1987, having doubled since the beginning of the 1980s, and generally declined through the 1990s after the fall of the USSR.

Health and Defense R&D at the end of the XX^{th} century. At the end of the 1990s, a major shift occurred with the doubling of the budget for medical research at the National Institutes for Health from 1998 to 2003. A third major boom in defense R&D was triggered by 9/11 in 2001 and lasted until the beginning of the Obama administration in 2008.

In summary, the narrative evidence discussed in this Appendix highlights that, especially compared to other types of government spending, public R&D was mostly driven by scientific, military and ideological goals, rather than by the endogenous policy response to the state of the U.S. economy. Accordingly, we propose to identify exogenous movements in public R&D using the short-run restrictions that while no macroeconomic variable can explain a large share of public R&D variation in the short-run (within the first year after the shock), public R&D is allowed (but not required) to have a significant impact on the economy in the short-run.

J Historical Decomposition of Public R&D based on the Military Spending Shocks

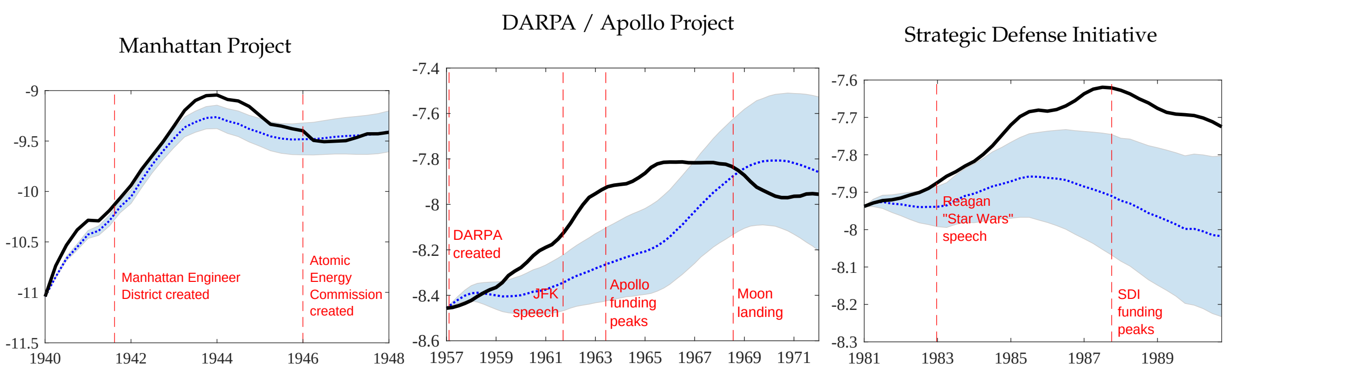
In this appendix, we perform a historical decomposition of the time series public of R&D (solid black lines) around three historical major events: (a) the Manhattan Project, (b) the DARPA / Apollo project, (c) the Strategic Defense Initiative. The blue lines and associated 68% central posterior bands in Figure J.1 represent the component of public R&D that can be explained by the military spending shock from a VAR(60) using military spending news, real public R&D per-capita, real GDP per-capita, real government spending per-capita, TFP, the short-term nominal interest rate, government deficit to GDP ratio and public debt to GDP ratio. The eight quarter moving-average of the time series of the military spending shocks (and associated 68% credible set) is plotted in the bottom panel.

The top panel of Figure J.1 reveals that the military spending shock explains a significant share of the public R&D increase around the Manhattan project, which occurred during World War II and was part of military R&D spending, but it can account only for a limited extent of the public R&D changes during other historical events that occurred in peacetime. The historical period associated with the DARPA/Apollo episode is a mixture of military and non-military spending, while the Reagan SDI increases are mostly defense spending that occurred, however, entirely during a peacetime period.

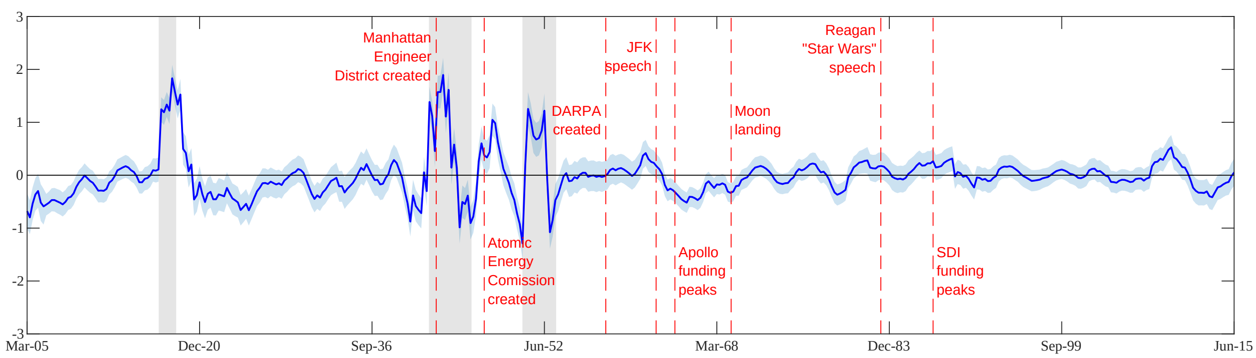
Finally, the time series of military spending shocks behind Figure J.1 and the time series of public R&D shocks behind Figure 9 have a correlation of only 0.17, suggesting that the two series are also driven by significantly different sources of variation.

Figure J.1: HISTORICAL ANALYSIS OF PUBLIC R&D AND MILITARY SPENDING SHOCKS

(a) Historical Decomposition of Public R&D Expenditure Around Key Events



(b) Time Series of Military News Shocks (eight quarter moving-average)



Notes: Panel (a) plots the historical decomposition of public R&D around three historical events: (i) the Manhattan project, (ii) DARPA and the Apollo program, (iii) the Strategic Defense Initiative (SDI). In each sub-panel, the solid black line is the historical surge in real per capita R&D spending by the government. The dotted blue line, and associated 68% posterior bands, show the part of the increase in R&D that can be explained by the effects of the *military spending shock*. Panel (b) plots a eight quarter moving-average of the military spending shock together with 68% posterior bands. Shaded areas represent major wars.

K The Effects of Public R&D Expenditure before and after 1948

As discussed in Section 2.3 and argued by Friedman (1952), identifying exogenous variation in government spending via military purchases by the U.S. government is attractive for at least two main reasons. First, wars are associated with large increases in public expenditure that are typically unrelated to the business-cycle, thereby ameliorating concerns about any possible reverse causality that may run from the state of the economy to government spending. Second, most wars the U.S. has been involved with have occurred outside the U.S. territory, thereby ameliorating concerns about any possible confounding effect that may come from the direct impact of the war itself. At the same time, this identification raises concerns about external validity, leaving the door open to the possibility that some of the effects that we have documented in the main text may not apply to other public spending categories.

In the main text, we have partially addressed the external validity concern by proposing a novel identification strategy that isolates exogenous variation in public R&D spending. This is also one of the very few categories of government expenditure that displays virtually no cyclicality at all. Still, while in Section 6.2 and Appendix I, we have discussed and presented evidence on several examples where large public R&D spending increases took place during peacetime, all major war episodes in the 125 years we consider were associated with significant spikes in public R&D. Furthermore, our long historical sample has witnessed several significant changes in economic policies and the structure of the U.S. economy: it is unclear the extent to which sub-sample instability may distort long-run inference.

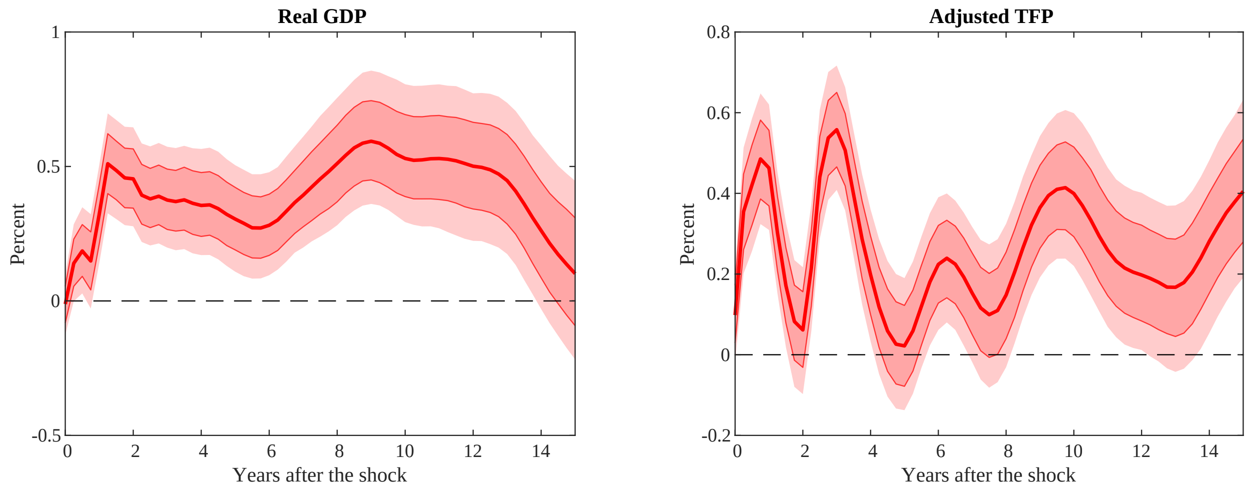
It is challenging to address sub-sample stability when dealing with inference over a 15-year horizon and 125 years of data, i.e. using 8 non-overlapping sub-samples of 15 years. It is particularly so for military spending, because the largest shocks are all concentrated in the first half of the sample. We have verified that dropping any 2 of these 8 sub-samples and using only 95 years of data at a time (i.e. over the samples 1890-1985, 1905-2000, 1920-2015) leads to very similar results as the baseline. Attempts to drop larger amounts of data and estimate a VAR over two sub-periods before and after WWII, however, lead to unstable results and huge confidence intervals. This is not surprising because, as mentioned above, the major wars occurred in the first half of the sample. This implies that, even when considering a pre-WWII period, one would need to include data that go at least until 1960 to be able to capture the full effects of the war. On the other hand, we obtain substantially more stable results below for R&D shocks, most likely because large R&D shocks are present in both halves of the sample.

Accordingly, we present impulse responses of output and productivity to a public R&D expenditure shock that has been identified (using the strategy in Section 6.2) and estimated (using a BVAR with sixty lags) over two sub-samples, namely before and after 1948. This is attractive for at least three main reasons. First, given the use of sixty lags in our BVAR, the effective post-WWII sample starts in 1963, and therefore it excludes all three major war episodes, namely WWI, WWII and the Korean war. Second, by comparing the IRFs of GDP and TFP in this Appendix across sub-samples as well as to those presented in Figure 10 of Section 6.2, we are able to provide prima faciae evidence on the influence that any possible sub-sample instability may exert on our baseline findings, which are based on a much longer historical sample. Third, starting from 1948, none of the quarterly time series in our dataset has been subject to any interpolation, and all of them come from readily available official sources.

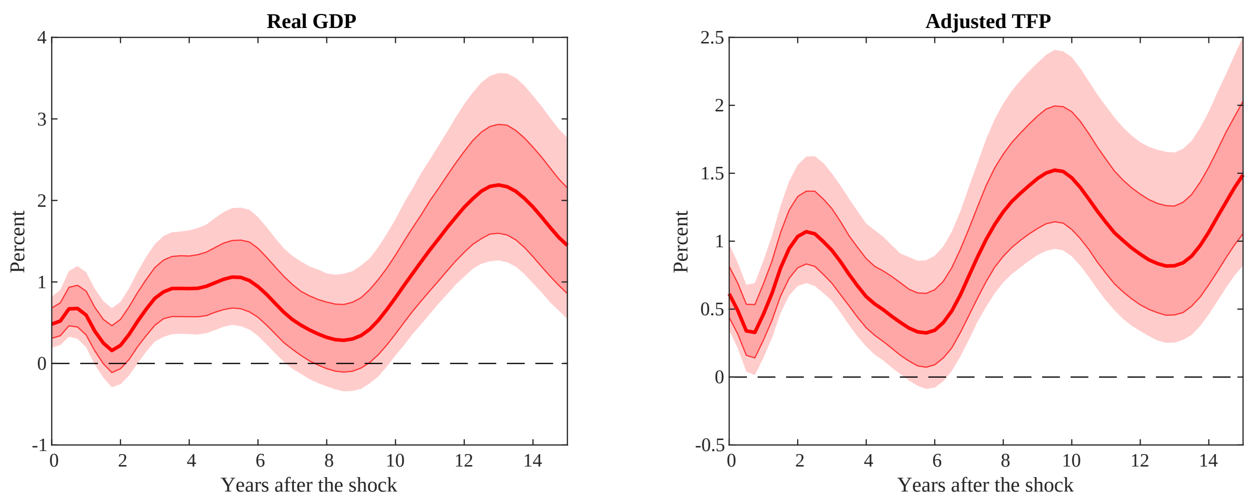
The results of this exercise are reported in Figure K.1 below. The left column refers to the impulse response of GDP while the right column summarizes the the impulse response of total factor productivity. Panel (a) exemplifies the effects of public R&D shocks in the pre-1948 period whereas the sample in panel (b) begins in 1948. As in the rest of the paper, solid lines stand for median posterior estimates while the shaded areas represent 68% and 90% posterior density intervals. Three main findings emerge from Figure K.1. First, the large, significant and beyond-business-cycle-frequency effects of an exogenous increase in government R&D expenditure on output and productivity are a robust feature of both the pre- and post-WWII eras. Second, the IRFs for output and productivity in the left column of Figure 10 over the full-sample appear to be a sort of average of their Figure K.1 counterparts across the two sub-periods. Third, the profile (over the forecast horizon) of the TFP response in the bottom right panel of Figure K.1 resembles closely the profile of the response of TFP to the non-defense public R&D shocks that Fieldhouse and Mertens (2023) estimate for the United States over the post-WWII period, using a very different identification strategy and very different data, based on government appropriations.

Figure K.1: IMPULSE RESPONSES TO PUBLIC R&D SHOCKS OVER DIFFERENT SUBSAMPLES

(a) 1890 Q1 - 1947 Q4



(b) 1948 Q1 - 2015 Q4



Notes: The impulse responses are based on an estimated VAR, using the two subsamples 1890-1947 and 1948-2015, with sixty lags of public R&D per capita, total government spending per capita, real GDP per capita, utilization-adjusted TFP, short-term interest rate, fiscal deficit to GDP ratio, government debt to GDP ratio, private investment per capita, total factor productivity, and patents. Public R&D, total government spending, GDP and TFP enter the VAR in log-levels. The public R&D shock is identified using the max-share method at the one-year forecast horizon. The size of the shock is normalized such as to increase government spending by 1 percent of GDP over the first year after the shock, based on a median G/Y ratio of 19%. The darker (lighter) shaded areas represent the central 68% (90%) high posterior density (HPD) intervals. The darker solid lines are the median estimates. Results are based on 5000 draws.

References

ANTOLIN-DIAZ, J., I. PETRELLA, AND J. F. RUBIO-RAMIREZ (2021): "Structural scenario analysis with SVARs," *Journal of Monetary Economics*, 117, 798–815.

BAI, J. (2004): "Estimating cross-section common stochastic trends in nonstationary panel data," *Journal of Econometrics*, 122, 137–183.

BEAUDRY, P., D. GALIZIA, AND F. PORTIER (2020): "Putting the Cycle Back into Business Cycle Analysis," *American Economic Review*, 110, 1–47.

BERGEAUD, A., G. CETTE, AND R. LECAT (2016a): "Long-Term Productivity Dataset," `http://longtermproductivity.com/download.html` (accessed March 2024).

——— (2016b): "Productivity Trends in Advanced Countries between 1890 and 2012," *Review of Income and Wealth*, 62, 420–444.

BUREAU OF THE CENSUS (1890-1929): "Statistical Abstract of the United States (1890-1929)," Washington, DC: U.S. Government Printing Office and Bureau of the Census, `https://www.census.gov/library/publications/time-series/statistical_abstracts.html` (accessed March 2024, data transcribed by authors).

——— (1949): "Historical Statistics of the United States, 1789–1945: A Supplement to the Statistical Abstract of the United States," U.S. Government Printing Office, `https://www.census.gov/library/publications/1949/compendia/hist_stats_1789-1945.html` (accessed March 2024, data transcribed by authors).

BUSH, V. (1954): "Science, the endless frontier," in *Science, the Endless Frontier*, Princeton University Press.

CHOW, G. C. AND A.-L. LIN (1971): "Best linear unbiased interpolation, distribution, and extrapolation of time series by related series," *The review of Economics and Statistics*, 372–375.

CHRISTIANO, L., M. EICHENBAUM, AND C. EVANS (2005): "Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy," *Journal of Political Economy*, 113, 1–45.

COMIN, D. AND M. GERTLER (2006): "Medium-Term Business Cycles," *American Economic Review*, 96, 523–551.

DUPREE, A. H. (1986): *Science in the federal government*, Johns Hopkins University Press.

FERNALD, J. G. (2012): "A Quarterly, Utilization-Adjusted Series on Total Factor Productivity," Federal Reserve Bank of San Francisco Working Paper Series 2012-19.

FIELDHOUSE, A. J. AND K. MERTENS (2023): "The Returns to Government R&D: Evidence from U.S. Appropriations Shocks," Working Papers 2305, Federal Reserve Bank of Dallas.

FRIEDMAN, M. (1952): "Price, Income, and Monetary Changes in Three Wartime Periods," *The American Economic Review*, 42, 612–625.

GIANNONE, D., M. LENZA, AND G. E. PRIMICERI (2015): "Prior selection for vector autoregressions," *Review of Economics and Statistics*, 97, 436–451.

GORDON, R. J. (2007): *The American business cycle: Continuity and change*, vol. 25, University of Chicago Press.

GÜRKAYNAK, R. S., B. SACK, AND E. SWANSON (2005): "Do Actions Speak Louder Than Words? The Response of Asset Prices to Monetary Policy Actions and Statements," *International Journal of Central Banking*, 1.

IMBS, J. M. (1999): "Technology, growth and the business cycle," *Journal of Monetary Economics*, 44, 65–80.

JORDÀ, Ò., M. SCHULARICK, AND A. M. TAYLOR (2017): "Macrohistory Database," https://www.macrohistory.net/database/ (accessed March 2024).

JORDÀ, Ò., S. R. SINGH, AND A. M. TAYLOR (2020): "The Long-Run Effects of Monetary Policy," Working Paper 26666, National Bureau of Economic Research.

MÜLLER, U. K. AND M. W. WATSON (2020): "Low-Frequency Analysis of Economic Time Series," Working Papers 2020-13, Princeton University. Economics Department.

NATIONAL BUREAU OF ECONOMIC RESEARCH (1986): "Tables from: *The American Business Cycle: Continuity and Change*, edited by Robert J. Gordon," Cambridge, MA: National Bureau of Economic Research, https://www.nber.org/research/data/tables-american-business-cycle (accessed March 2024).

——— (1997): "NBER Macrohistory Database," Cambridge, MA: National Bureau of Economic Research, https://www.nber.org/research/data/nber-macrohistory-database (accessed March 2024.

NATIONAL RESEARCH COUNCIL (1995): *Allocating federal funds for science and technology*, National Academies Press.

RAMEY, V. A. (2011): "Identifying Government Spending Shocks: It's all in the Timing," *Quarterly Journal of Economics*, 126, 1–50.

RAMEY, V. A. AND S. ZUBAIRY (2018a): "Government Spending Multipliers in Good Times and in Bad: Evidence from US Historical Data," *Journal of Political Economy*, 126, 850–901.

——— (2018b): "Replication data for: "Government Spending Multipliers in Good Times and in Bad: Evidence from US Historical Data"," Chicago, IL: University of Chicago Press, https://www.journals.uchicago.edu/doi/suppl/10.1086/696277.

ROMER, C. D. AND D. H. ROMER (2004): "A New Measure of Monetary Shocks: Derivation and Implications," *American Economic Review*, 94, 1055–1084.

U.S. BUREAU OF ECONOMIC ANALYSIS ([YYYY]): "Unpublished Annual Estimates of U.S. Investment, 1901–present," Unpublished data provided by the BEA upon request (accessed March 2024).

WELCH, I. AND A. GOYAL (2008a): "A comprehensive look at the empirical performance of equity premium prediction," *The Review of Financial Studies*, 21, 1455–1508.

——— (2008b): "Data for: "A Comprehensive Look at the Empirical Performance of Equity Premium Prediction"," Data available at Amit Goyal's personal website: https://sites.google.com/view/agoyal145 (accessed March 2024).