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.

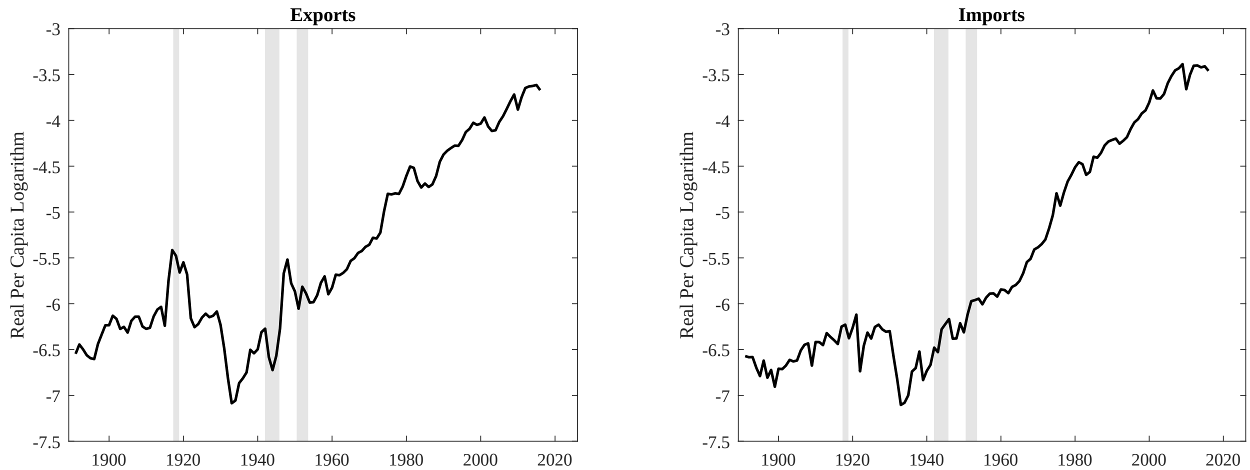
B Data construction

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.

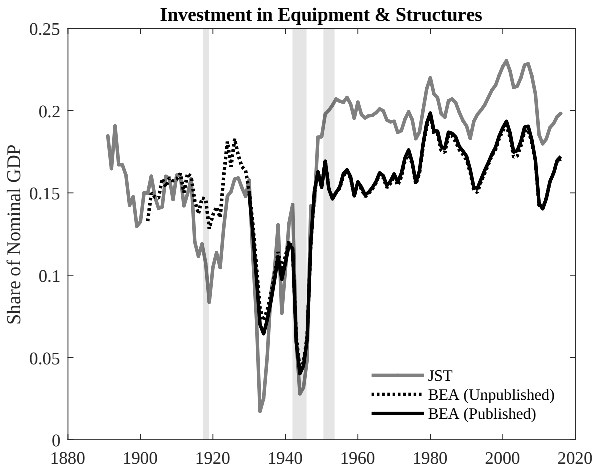
quarterly real GDP from Ramey and Zubairy (2018a) (Ramey and Zubairy, 2018b) to perform the interpolation. The resulting series is visible in the middle panel of Figure B.3.

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:

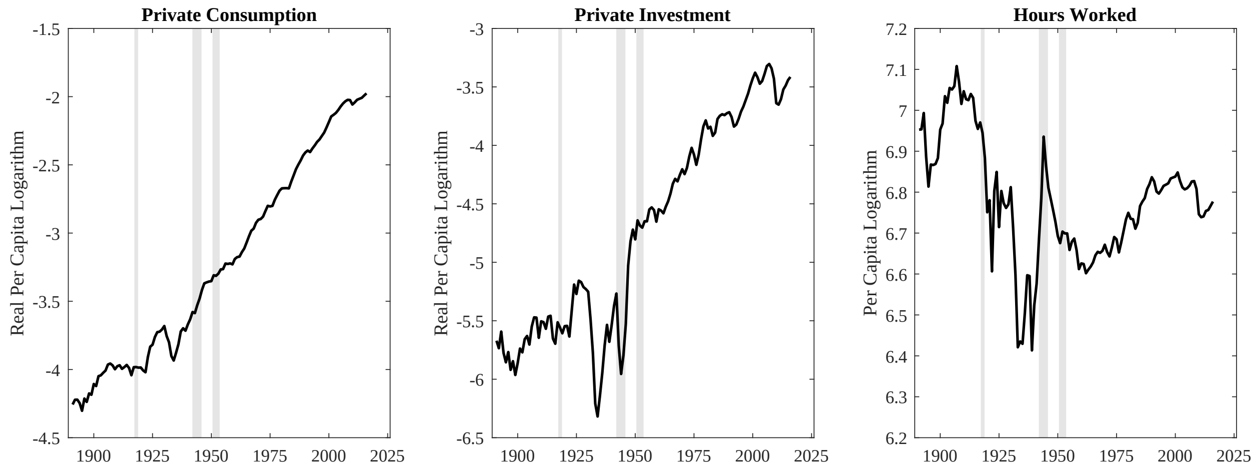
$$c_t^{JST} - \bar{c}_t^{JST} + \bar{c}_t^{NA} \tag{2}$$

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

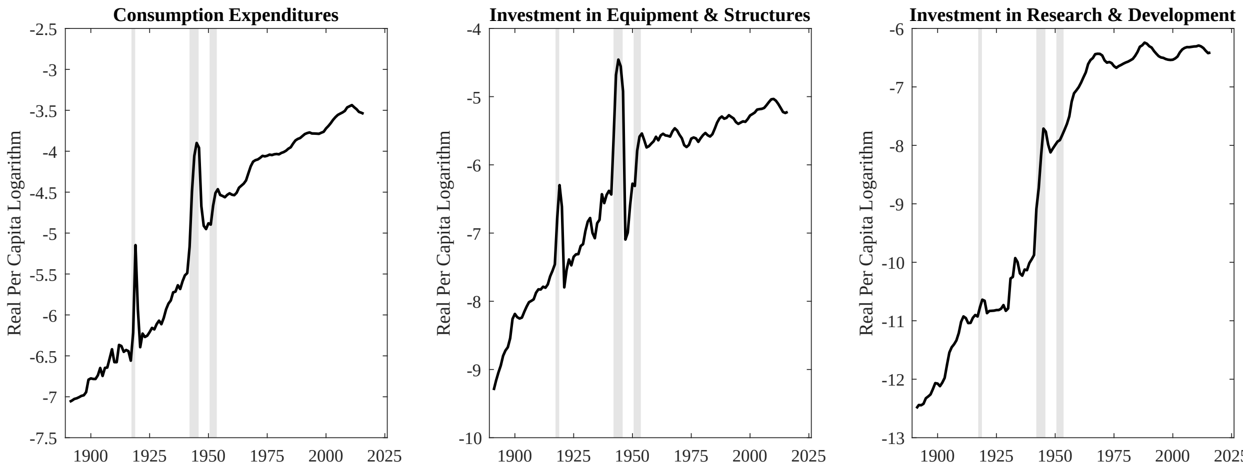


Notes. Private consumption and investment are deflated using the GDP deflator. All variables are scaled by the civilian population. Shaded areas represent the three major wars in the sample: World War I, World War II, and the Korean War.

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.

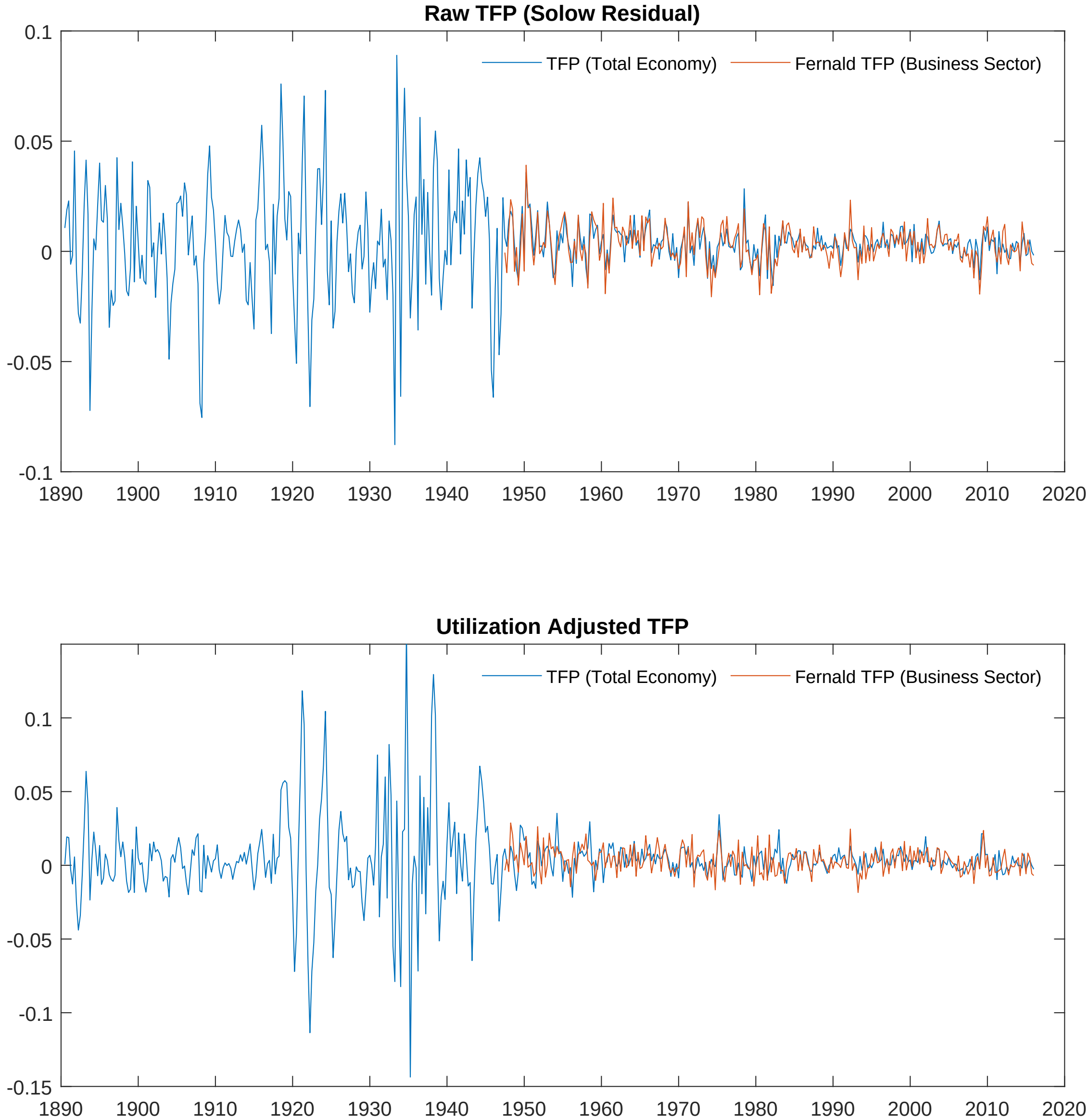
Finally, the quarterly time series for Total Factor Productivity (TFP) has been constructed in two steps. First, we obtain annual measures of hours worked and the capital stock from Bergeaud et al. (2016b) (Bergeaud et al., 2016a).[1] These annual time series are interpolated to quarterly frequency. 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 $\lambda = 1600$.

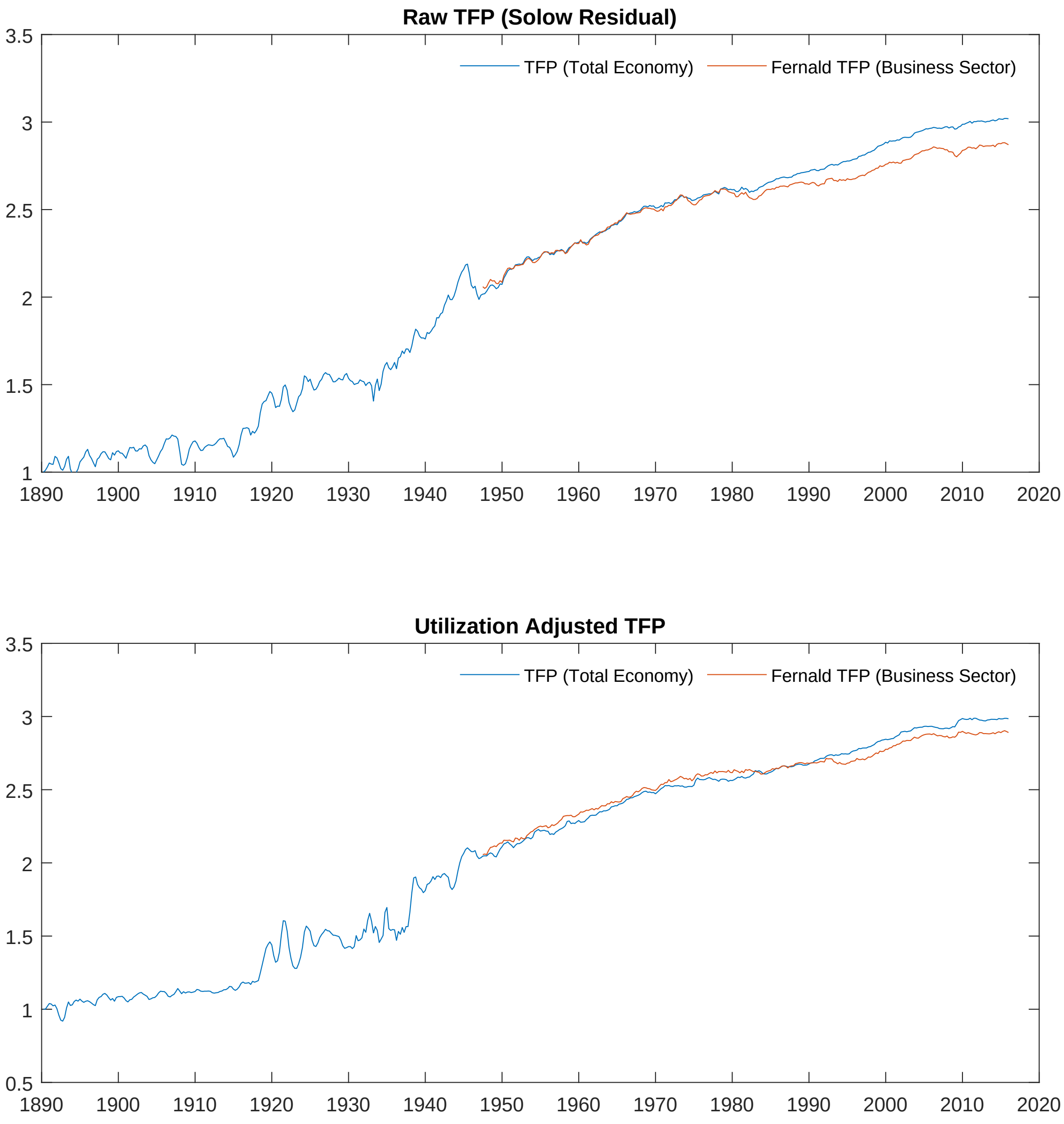
As shown in Figure B.5, which displays growth rates, and Figure B.6, which depicts log-levels, our historical quarterly time series of adjusted TFP, which refers to the whole economy, moves very closely to the more sophisticated and more data intensive measure proposed by Fernald (2012), which covers 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



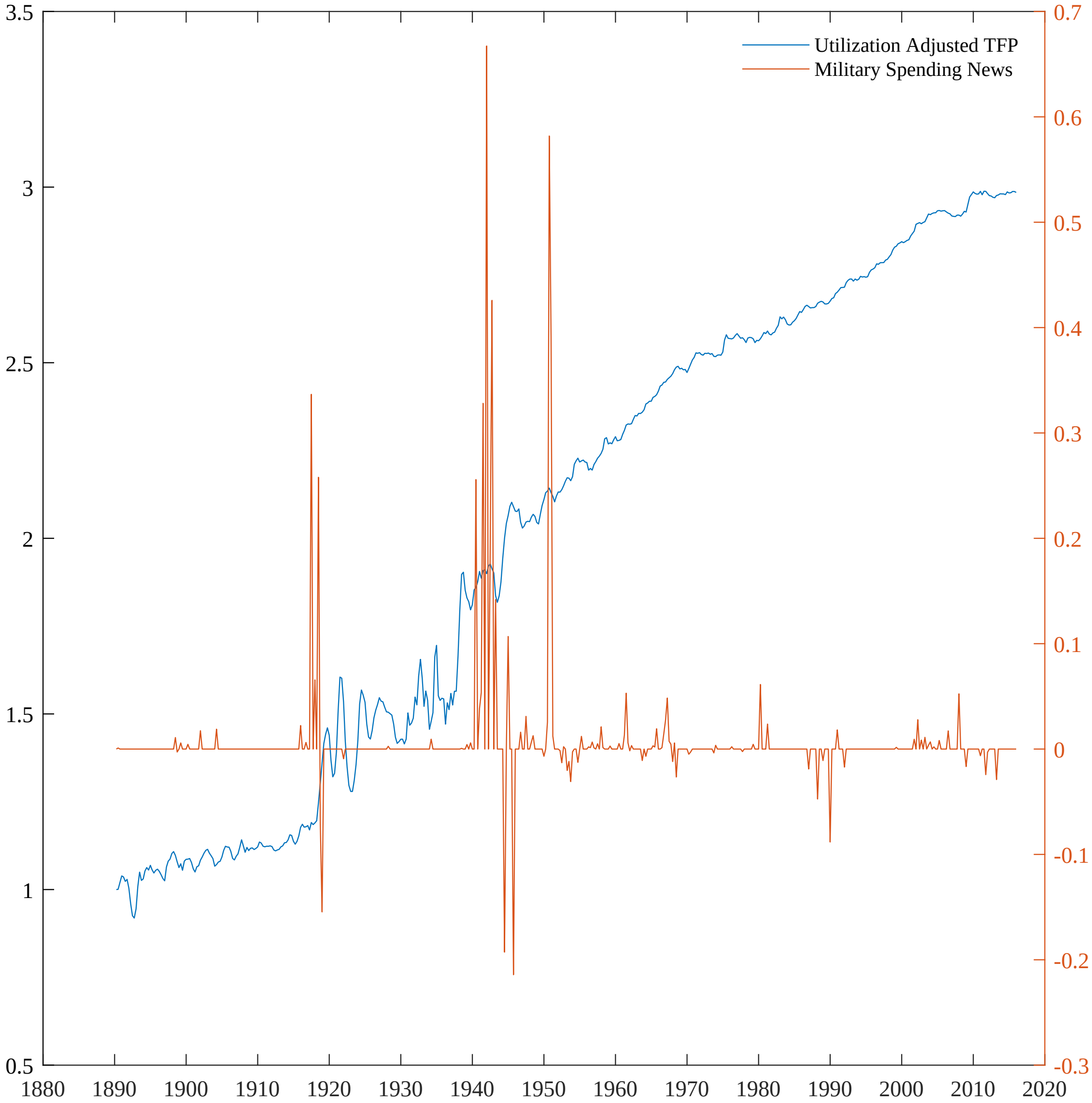
Notes. TFP Growth Rates as described in the Text. Top (bottom) row refers to the raw (utilization adjusted) TFP series.

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

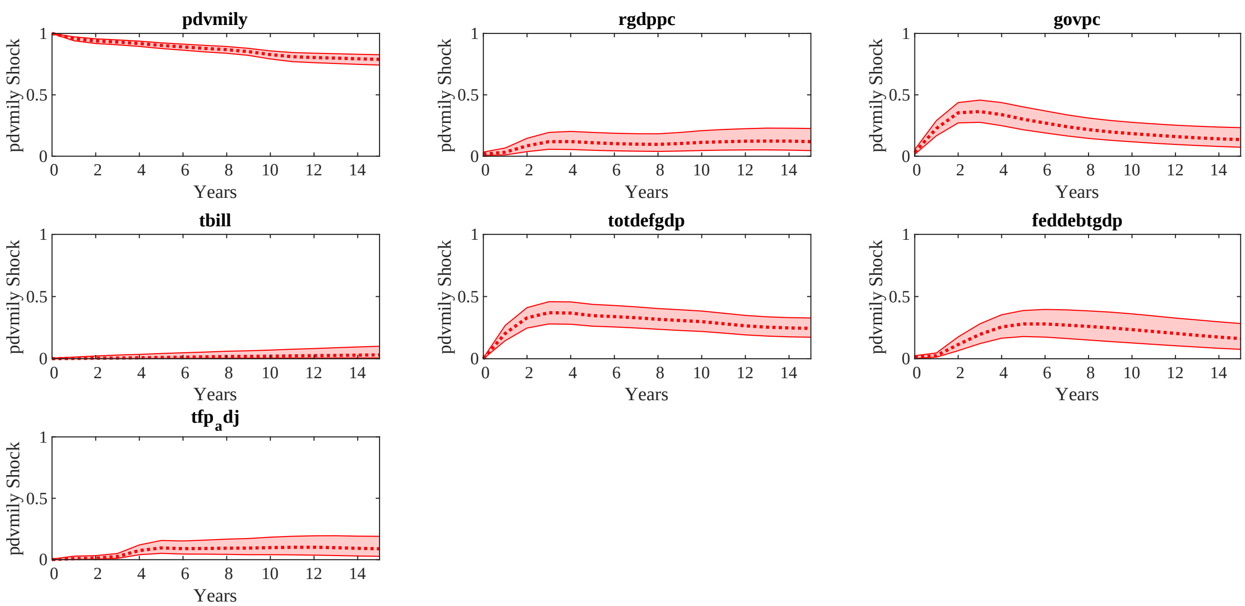


Notes. Utilization-adjusted TFP levels as described in the Text. The military spending news as a percentage of GDP (right axis) is from Ramey and Zubairy (2018a) (Ramey and Zubairy, 2018b)

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

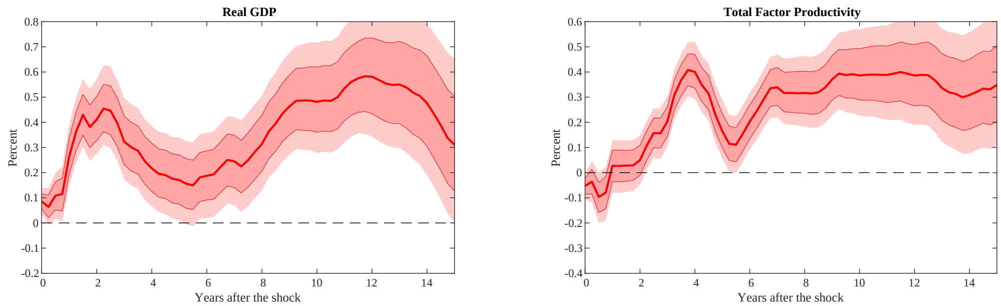
In this section, we present the impulse responses of output and productivity based on variants of the VAR(60) estimated in the main text. The four versions differ in the tightness of the prior hyper-parameters λ (which controls the tightness of the "Minnesota" prior) and θ (which controls the tightness 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.

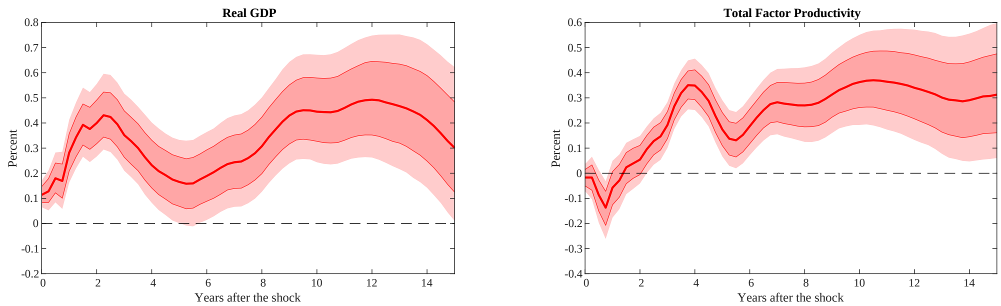
Three main results emerge from this sensitivity analysis. First, progressively tighter priors on the λ hyperparameter of the VAR(60), which are visible going down the rows of Figure D.1, are associated both with smoother shapes of the impulse responses and also with progressively smaller effects at long 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

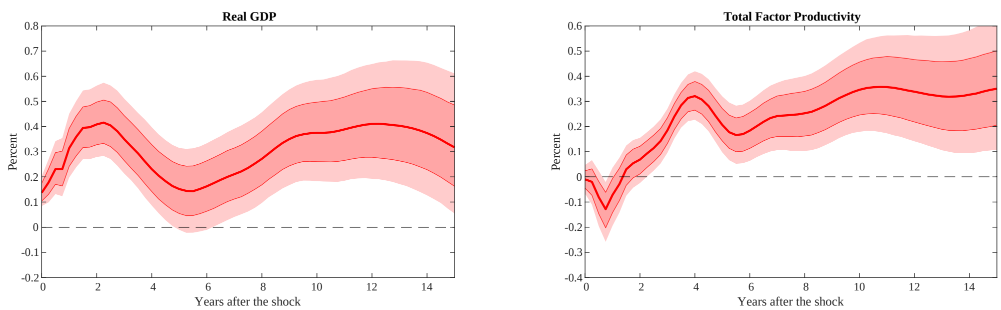
(a) $\lambda = 1$



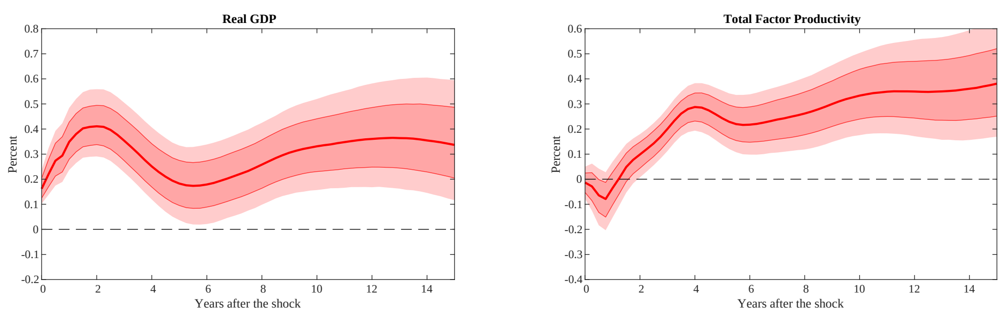
(b) $\lambda = 0.4$



(c) $\lambda = 0.2$



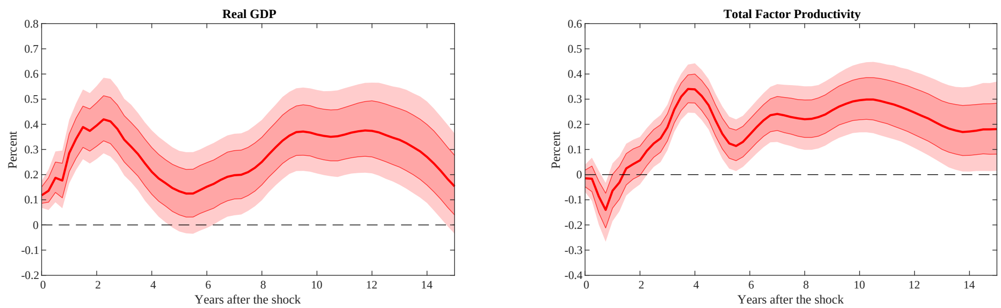
(d) $\lambda = 0.1$



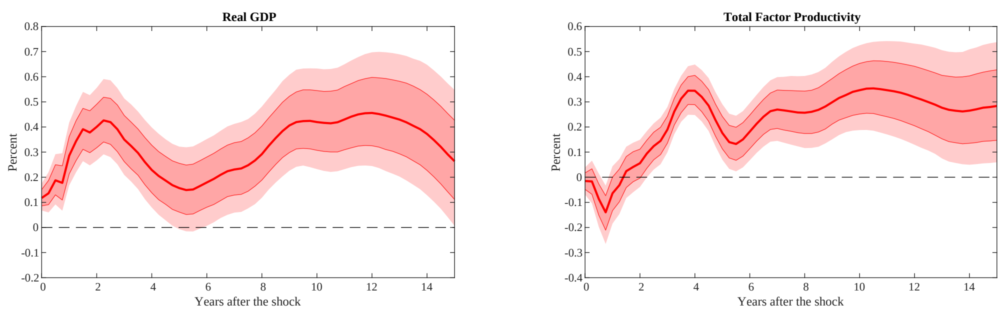
Notes: The solid lines represent the median posterior response. The darker (lighter) shadow area represents the 68^{th} (90^{th}) HPD interval. All specifications use sixty lags of military spending news, government spending, GDP, GDP deflator, treasury bills, deficit to GDP ratio and debt to GDP ratio 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

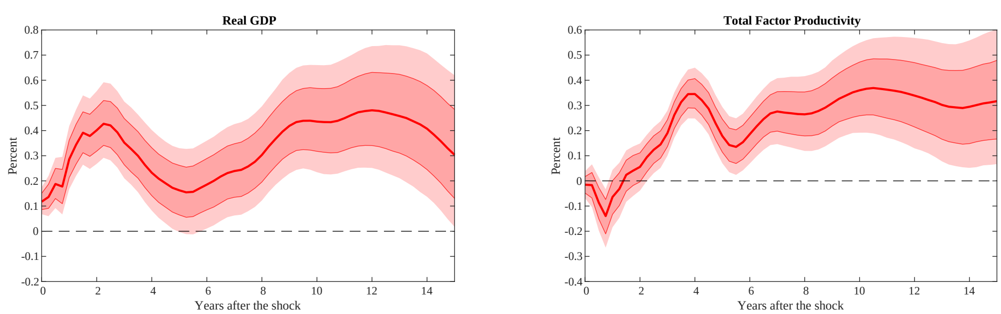
(a) $\theta = 1$



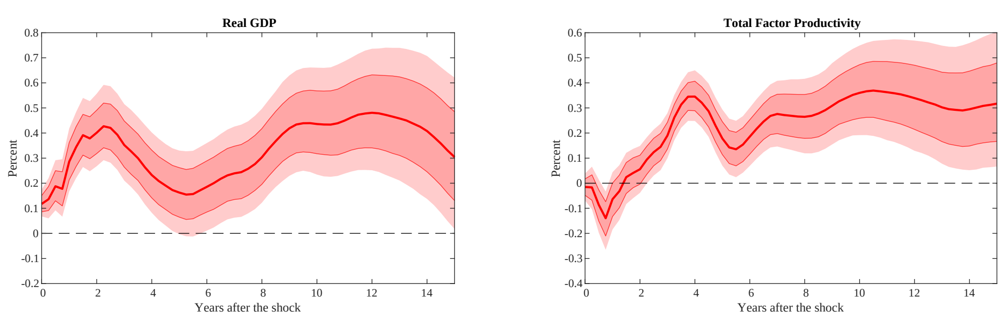
(b) $\theta = 0.1$



(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

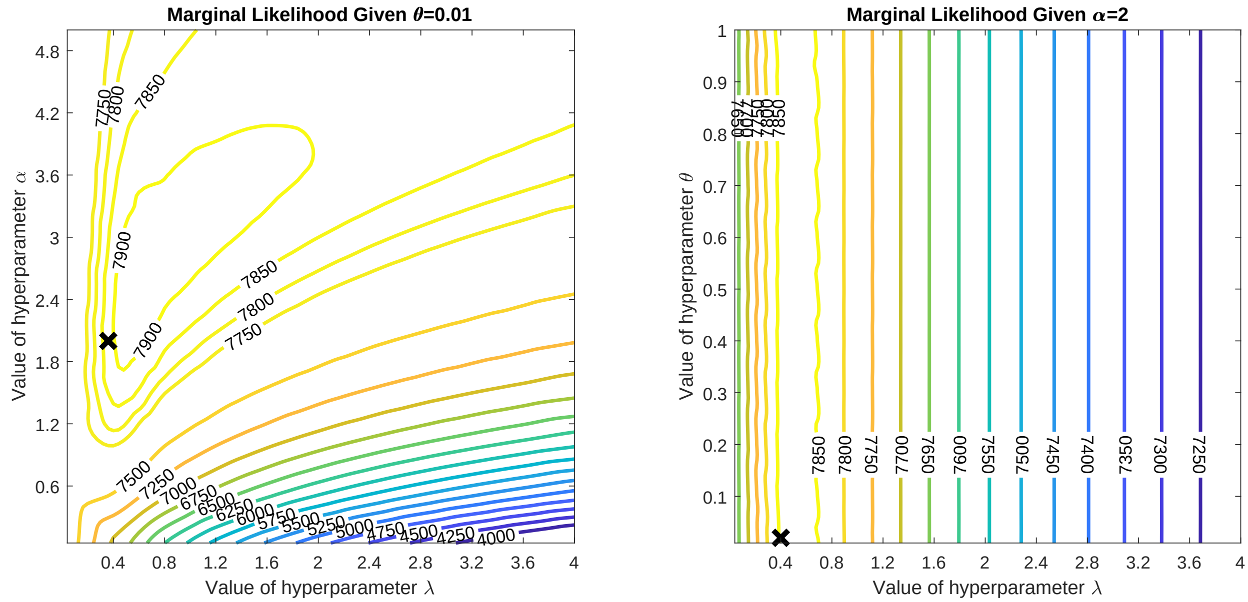
In this section we provide further details on the marginal likelihood of the model and the estimation with hierarchical priors. Our approach follows Giannone et al. (2015) and uses exactly their settings, maximization algorithm, and hyperparameter choices.

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.

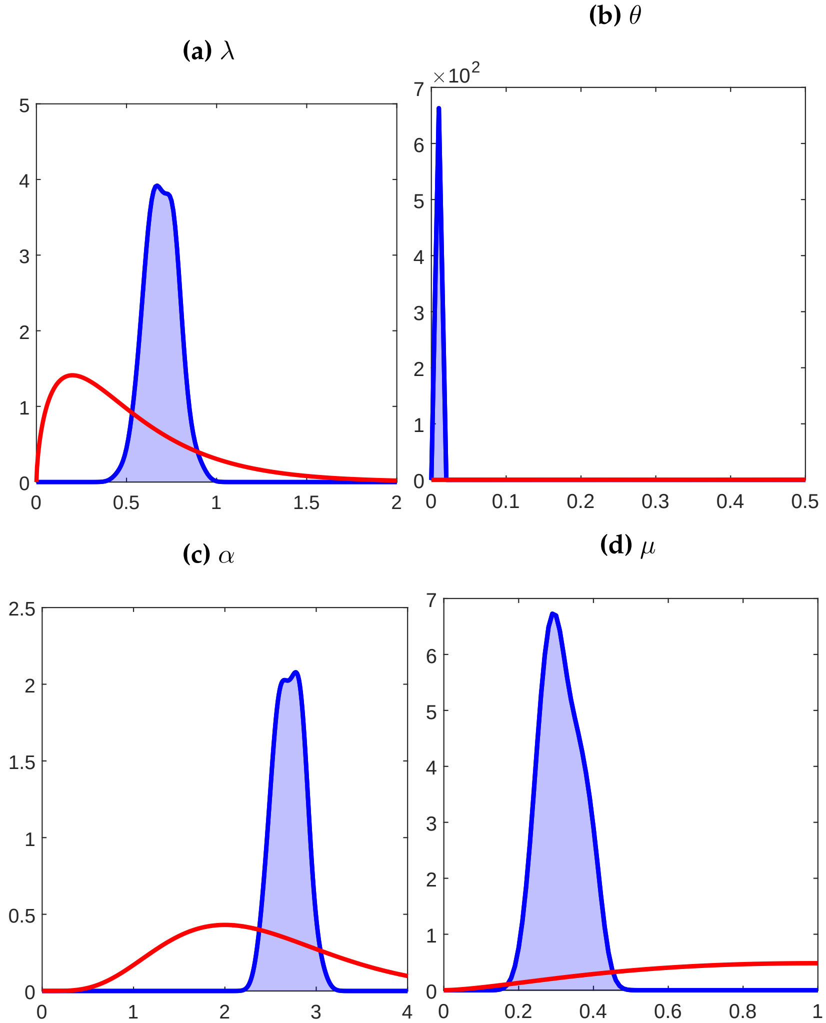
Figure E.1: Marginal Likelihood as a Function of Prior Hyperparameters



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.

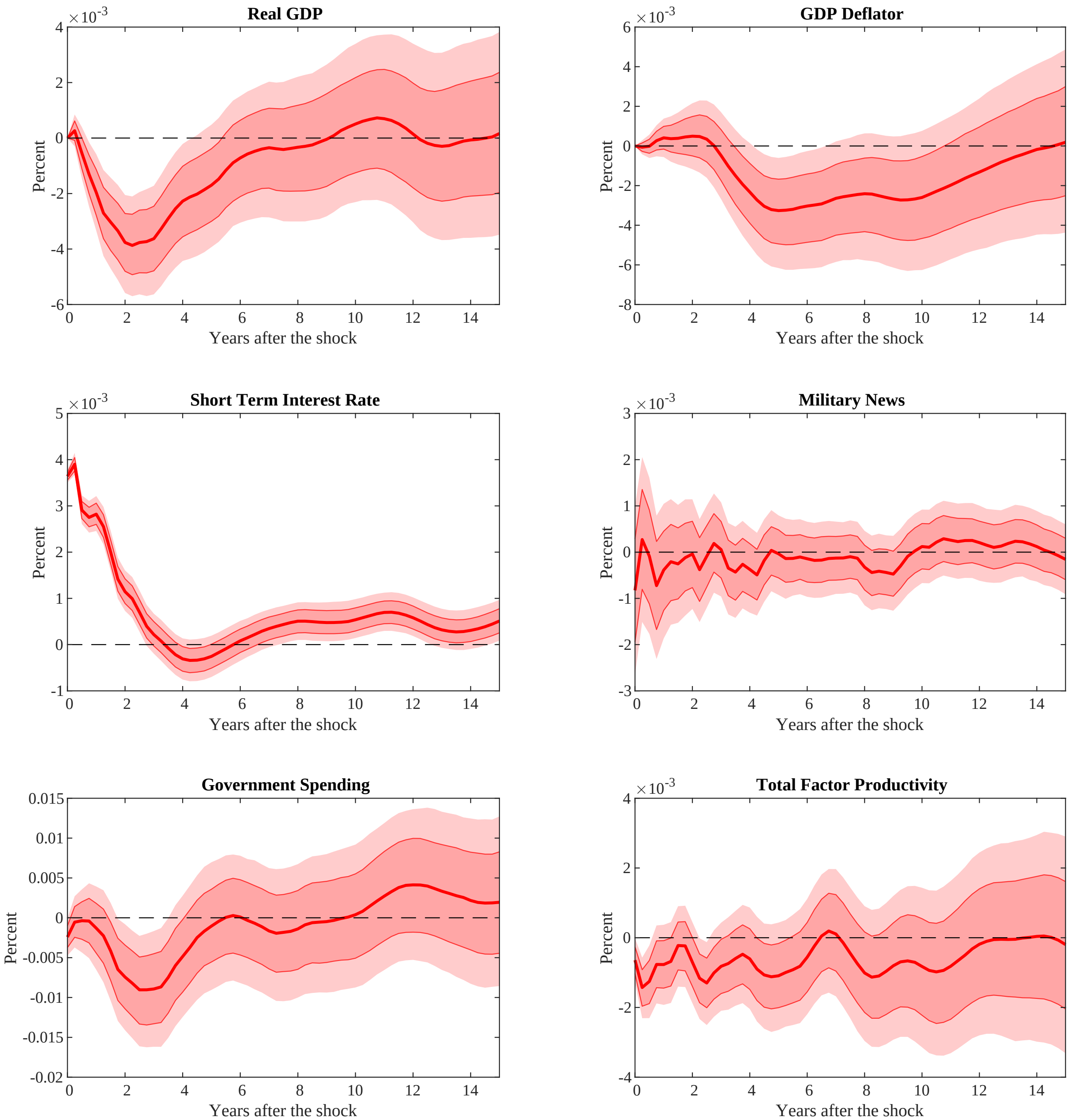
To evaluate this hypothesis, we augment our baseline VAR(60) with the GDP deflator and identify the monetary policy shock via a Cholesky factorization in which real GDP and the GDP deflator are ordered before the short-term interest rate. The idea behind this identification, which has a long tradition 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.

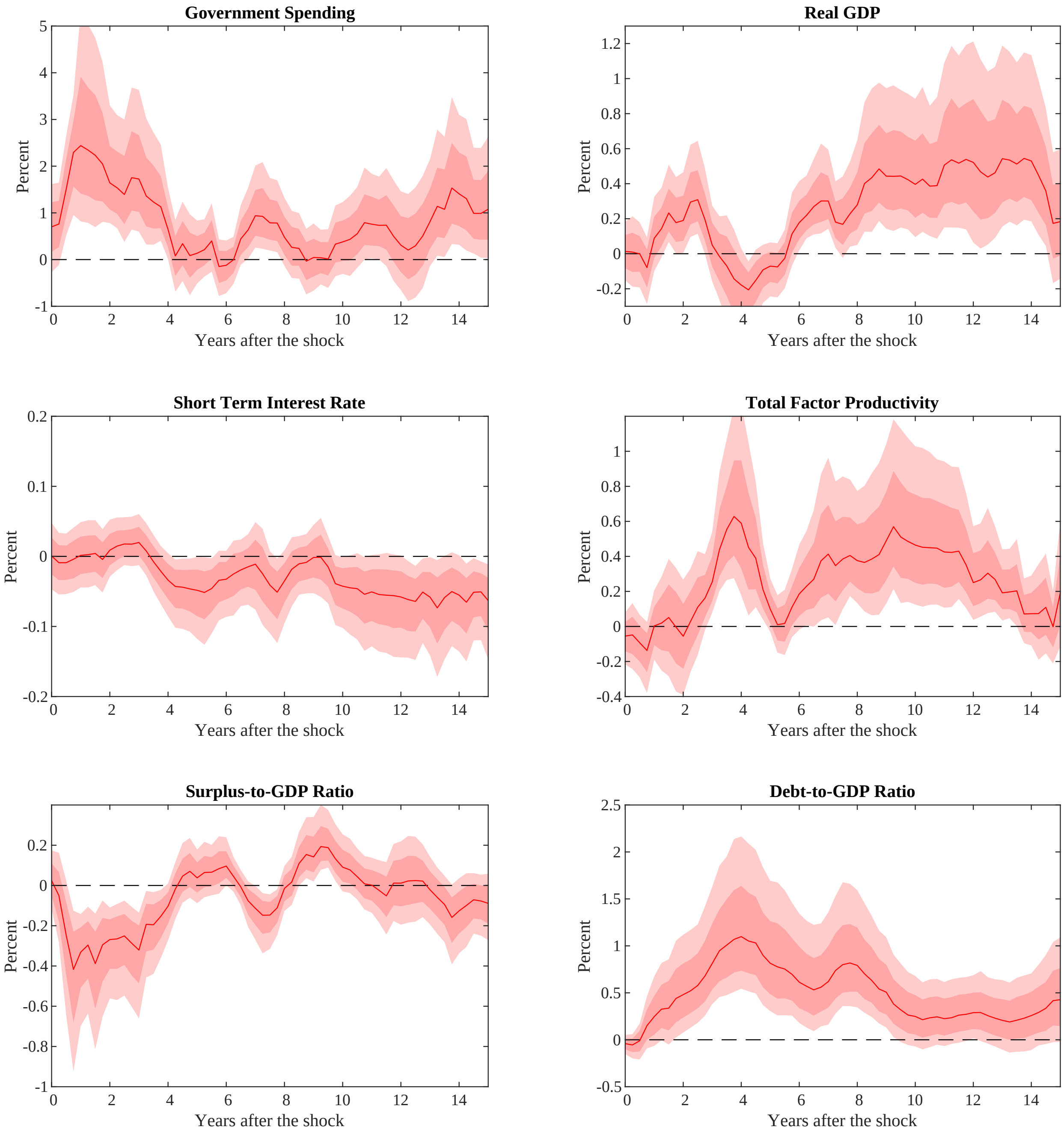
G Local Projections with Principal Components

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

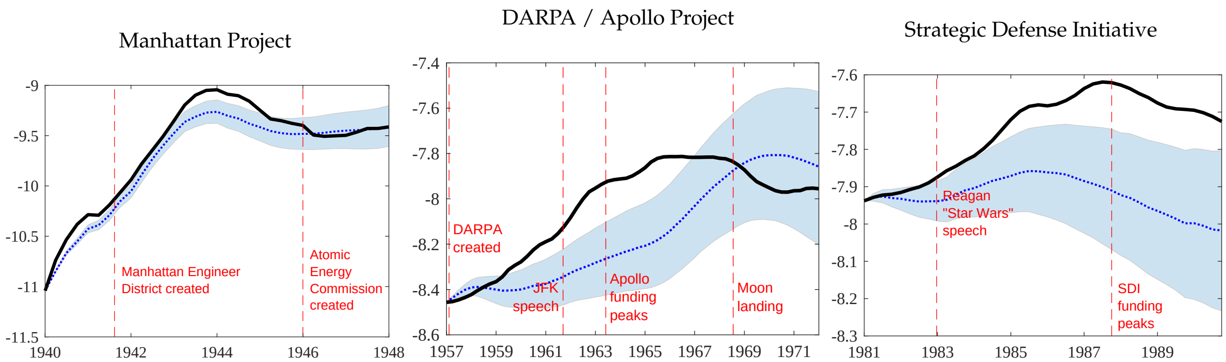
In this appendix, we perform a historical decomposition of the time series public of R&D (solid black lines) around three historical major events: (a) the Manhattan Project, (b) the DARPA / Apollo project, (c) the Strategic Defense Initiative. The blue lines and associated 68% central posterior bands 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.

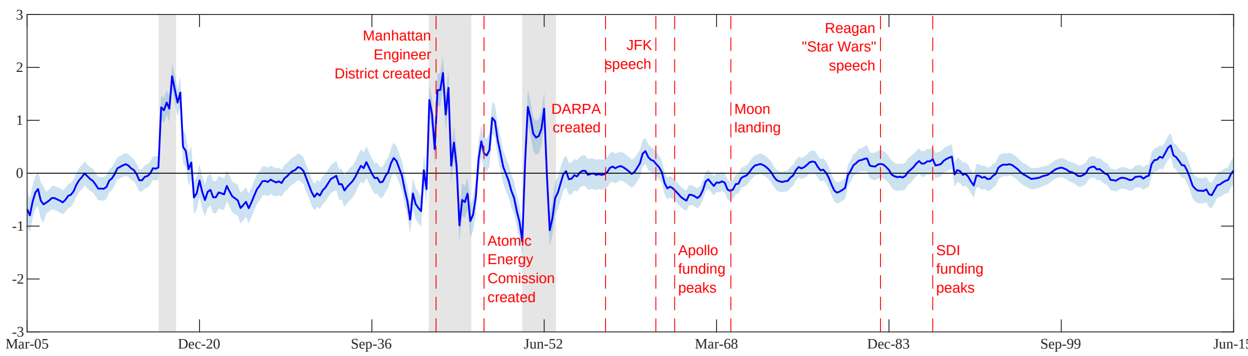
Finally, the time series of military spending shocks behind Figure J.1 and the time series of public R&D shocks behind Figure 9 have a correlation of only 0.17, suggesting that the two series are also driven by significantly different sources of variation.

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.

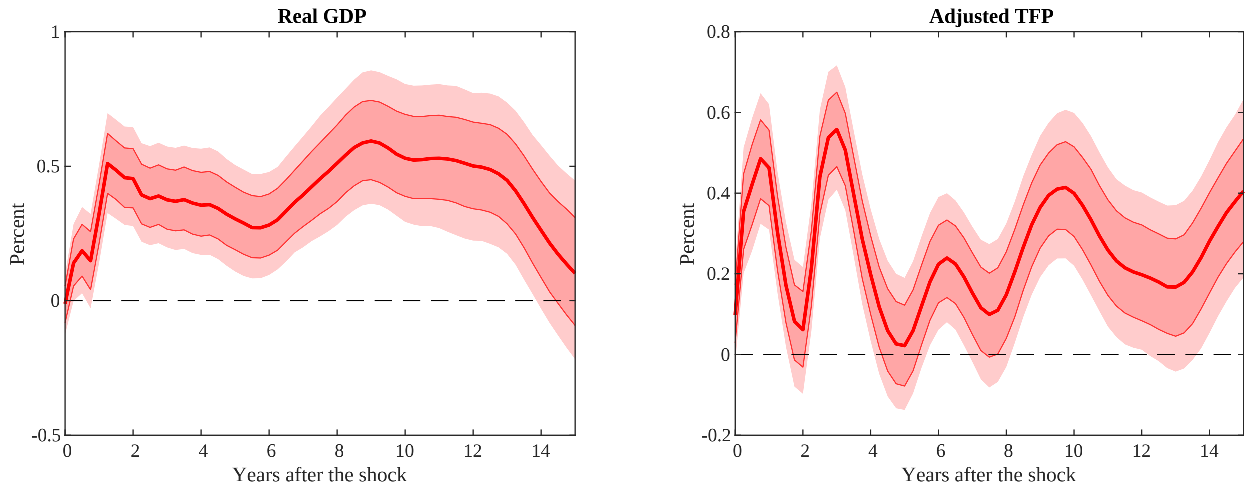
It is challenging to address sub-sample stability when dealing with inference over a 15-year horizon and 125 years of data, i.e. using 8 non-overlapping sub-samples of 15 years. It is particularly so for military spending, because the largest shocks are all concentrated in the first half of the sample. 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.

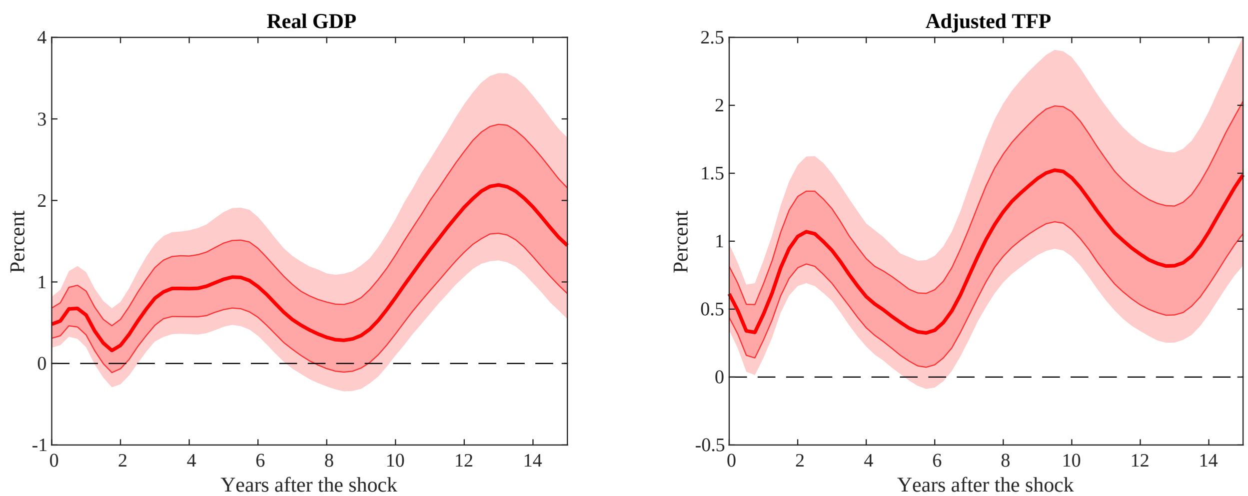
The results of this exercise are reported in Figure K.1 below. The left column refers to the impulse response of GDP while the right column summarizes the the impulse response of total factor productivity. Panel (a) exemplifies the effects of public R&D shocks in the pre-1948 period whereas the sample 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).