

The Long-Run Effects of Government Spending*

Juan Antolin-Diaz

Paolo Surico

London Business School

London Business School and CEPR

July 2022

Abstract

We study the dynamic effects of military spending and other government outlays using 125 years of U.S. quarterly data and Vector Auto Regressions (VAR) with up to sixty lags. The output multiplier is below one at business-cycle frequencies but is above one in the long-run. A public spending expansion of 1% of GDP leads to a sustained increase in inflation of 0.15% per year. Military spending crowds out private investment over the first four years but crowds it in at longer horizons. Over the medium-term, innovation and productivity rise persistently, leading to a second output expansion. We show that these dynamics are most likely driven by government R&D expenditure rather than by public investment or public consumption, highlighting an important channel through which fiscal policy can support economic growth beyond the business-cycle.

Keywords: government R&D, long-run, TFP, innovation, output multiplier, inflation.

JEL Classification Codes: E31, E52, E62.

*We are grateful to Michele Andreoli, Tim Besley, Olivier Blanchard, Luca Gambetti, Ethan Ilzetzi, Diego Känzig, Elias Papaioannou, Ivan Petrella, Mikkel Plagborg-Møller, Franck Portier, Giovanni Ricco, Juan Rubio-Ramirez, Christian Wolf and participants at the Barcelona Summer Forum for very useful comments and suggestions. Paolo Surico gratefully acknowledges financial support from the European Research Council (Grant Agreement No. 771976). Correspondence: London Business School, Regent's Park, NW1 4SA London, United Kingdom. Email: jantolindiaz@london.edu and psurico@london.edu.

1 Introduction

The financial crisis of 2007-09 and the global pandemic of 2020-22 have triggered a wave of government interventions around the world that, by many historical standards, have been very significant in magnitude and breadth. In response to that, policy and academic research has produced an extraordinary number of empirical studies that have considerably moved forward the frontier of our knowledge about the transmission of fiscal policy.

Despite these important advances, most time-series analyses have focused on business-cycle frequencies and little is known about the long-run effects of government spending. This is particularly surprising in the light of the historical case studies from an influential empirical literature showing that large public expenditures (e.g. on roads, railroads and military build-ups) have led to sizable increases in productivity across U.S. industries, counties and plants ([Fernald, 1999](#); [Donaldson and Hornbeck, 2016](#); [Ilzetzki, 2022](#)).

In this paper, we use 125 years of quarterly data for the U.S. and time-series models with a rich lag structure to uncover novel evidence on the macroeconomic effects of government spending on output, prices, productivity and innovation in the long-run. We show that the combination of long historical data and a high number of lags is crucial for the ability of popular macro models, such as Vector Autoregressions (VAR) and Local Projections (LP), to identify aggregate effects beyond the business-cycle, traditionally defined as frequencies between six quarters and eight years.

Main results. The output multiplier is above one on impact, but turns to values significantly smaller than one about a year after the shock. The increase in public spending has negative short-run effects on investment and consumption while triggering a surge in government R&D, deficit and public debt. Prices rise persistently

for three years.

Between four and eight years after the shock, government spending falls below pre-shock levels. This halts the output expansion and leads to a fiscal surplus that counterbalances the rise in public debt. The responses of private investment, innovation and total factor productivity become significantly positive, whereas the price level reverts its course.

In the long-run, the surge in investment, innovation and total factor productivity persists, leading to a second output expansion. As the change in public spending is no longer statistically significant, the dynamic effects on GDP imply a long-run multiplier above one. Eventually, the debt-to GDP ratio and the price index return to their pre-shock levels.

Our empirical framework is a quarterly Bayesian VAR with sixty lags and moderate shrinkage via the prior distributions ([Giannone et al., 2015](#)). These choices fulfill our desire to balance the bias-variance trade-off discussed by [Li et al. \(2021\)](#), while retaining the ability to identify any possible long-run effect. Once we allow for a generous lag structure and shrinkage, however, Bayesian LP and VARs produce very similar results. In contrast, popular specifications in empirical macro, such as a VAR(4) or a LP with four lags of all relevant variables as controls, tend to estimate small and insignificant long-run dynamics.

Our main results are based on the military spending news built by [Ramey and Shapiro \(1998\)](#), [Ramey \(2011b\)](#) and [Ramey and Zubairy \(2018\)](#) but are robust to using total government spending and the identification in [Blanchard and Perotti \(2002\)](#). Furthermore, we show that an exogenous increase in public R&D produces similar long-run dynamics on productivity and GDP. However, this is much less (or not at all) the case for a shock to government investment (consumption). Finally, excluding large war episodes, such as WWI, WWII or the Korean war, or making alternative assumptions on the GDP trend or the prior distributions do not overturn our findings.

Contribution. We seek to make several advances relative to the literature discussed below. First, we document a novel dimension of heterogeneity in the effects of government spending on output, namely across forecast horizons, which can reconcile seemingly conflicting estimates on the size of the output multiplier in earlier macro studies. Second, we uncover an important role for the corporate sector to shape the long-run effects of fiscal policy: the responses of investment, innovation and productivity to a government spending shock lead the responses of output at horizons beyond four years, even though the changes in government spending are no longer significant. Third, we present suggestive evidence that the long-run effects of military spending are driven by public R&D rather than by government consumption, and that a surge in public R&D and, to a lesser extent, public investment can have long-run effects on output even when not triggered by military spending. Fourth, our findings are based on a long historical dataset for which we have compiled detailed archival information to construct new quarterly series of U.S. private investment, total factor productivity, patents, public R&D, government investment and government consumption since 1889Q1. Finally, we show that the prevailing practice in the empirical macro literature of using short lag structures may suffer from truncation and omitted variable biases, which make them less suited to identify long-run effects.

Related Literature. A voluminous empirical literature has studied the macroeconomic effects of government spending on 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 short-run government spending multiplier lies between 0.6 and 1.5,

across the reviewed papers. Our focus on the long-run is a distinctive feature relative to earlier studies.

Another strand of research seeks to reconcile the different estimates available in the literature by looking at potential sources of heterogeneity. These include trade openness and public debt ([Ilzetzki et al., 2013](#)), anticipation effects ([Ramey, 2011b](#); [Forni and Gambetti, 2016](#)), identification strategies ([Caldara and Kamps, 2017](#)), business cycle phases ([Auerbach and Gorodnichenko, 2012](#); [Caggiano et al., 2015](#)), zero lower bound on interest rates and economic slack ([Ramey and Zubairy, 2018](#)), fiscal expansions versus contractions ([Barnichon et al., 2022](#)) and the monetary policy regime ([Ascari et al., 2021](#)). Relative to these, we highlight that the size of the government spending multiplier varies with the forecast horizon, in a way that spans the range of estimates in earlier work.

An important literature focuses on the link between defense expenditure, R&D and productivity. For instance, [Moretti et al. \(2019\)](#) and [Deleidi and Mazzucato \(2021\)](#) find that military spending fosters innovation in the private sector while [Gross and Sampat \(2020\)](#), [Diebolt and Pellier \(2022\)](#) and [Ilzetzki \(2022\)](#) document that the two World Wars had long-lasting effects on U.S. patents and productivity. Our historical analysis extends these earlier findings to a much longer horizon and a different identification strategy by showing that public R&D can stimulate GDP growth beyond business-cycle frequencies.

Our results also speak to the public infrastructure research surveyed by [Ramey \(2020\)](#). For instance, [Fernald \(1999\)](#) and [Leff Yaffe \(2020\)](#) find that the U.S. interstate highway programme boosted industry-level productivity, while [Donaldson and Hornbeck \(2016\)](#) and [Hornbeck and Rotemberg \(2021\)](#) estimate that the U.S. national railroad network improved market access. We complement these studies by showing that public investment in equipment and infrastructure tends to have smaller long-run effects on productivity and GDP than public R&D spending.

A growing literature, surveyed by [Cerra et al. \(2022\)](#), studies the long-run effects of demand shocks. [Comin and Gertler \(2006\)](#) and [Beaudry et al. \(2020\)](#) lay out models with strong internal propagation mechanisms in which non-technology shocks have effects beyond business cycle frequencies. [Benigno and Fornaro \(2017\)](#) focus on stagnation traps triggered by weak aggregate demand. [Jordà et al. \(2020\)](#) 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 across U.S. states. [Cloyne et al. \(2022\)](#) estimate the long-run response of R&D, productivity and GDP to corporate and personal tax changes. We complement these analyses by offering an unprecedented evaluation of the long-run effects of military spending and other government outlays.

Structure of the paper. In Section 2, we present the VAR and LP specifications, the historical data and the identification strategy. The main findings on output, prices, the fiscal multiplier and the transmission mechanism via investment, productivity and innovation are reported in Section 3. In Section 4, we assess the role of public R&D, public investment and public consumption in shaping the responses of output and productivity at longer horizons. In Section 5, we investigate the role of lag length selection. The robustness of our results to a wide range of sensitivity exercises is the focus of Section 6. Conclusions are discussed in Section 7. In the Appendix, we provide details on the estimation and present further analyses.

2 Empirical framework

In this section, we motivate the empirical model and the estimation strategy 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 military spending news proposed 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 data with extended series for business investment, productivity, innovation, consumption, government investment and government consumption.

2.1 Model specification and estimation

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

$$\mathbf{y}'_t \mathbf{A}_0 = \sum_{\ell=1}^p \mathbf{y}'_{t-\ell} \mathbf{A}_\ell + \mathbf{c} + \boldsymbol{\varepsilon}'_t \quad \text{for } 1 \leq t \leq T \quad (1)$$

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}_ℓ 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 $1 \times n$, the letter p refers to the lag length, whereas T denotes the sample size. The vector $\boldsymbol{\varepsilon}_t$, conditional on past information and the initial conditions $\mathbf{y}_0, \dots, \mathbf{y}_{1-p}$, is Gaussian with zero mean and covariance matrix \mathbf{I}_n , the $n \times n$ identity matrix.

Denoting $\mathbf{A}'_+ \equiv [\mathbf{A}'_1 \cdots \mathbf{A}'_p \ \mathbf{c}']$, 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 $\mathbf{x}'_t = [\mathbf{y}'_{t-1}, \dots, \mathbf{y}'_{t-p}, 1]$, $\mathbf{B} = \mathbf{A}_+ \mathbf{A}_0^{-1}$, $\mathbf{d} = \mathbf{c} \mathbf{A}_0^{-1}$, $\mathbf{u}'_t = \boldsymbol{\varepsilon}'_t \mathbf{A}_0^{-1}$, and $\mathbb{E}[\mathbf{u}_t \mathbf{u}'_t] = \boldsymbol{\Sigma} = (\mathbf{A}_0 \mathbf{A}'_0)^{-1}$. 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.

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, $\boldsymbol{\Theta} = (\mathbf{A}_0, \mathbf{A}_+)$.¹

¹For instance, given a value $\boldsymbol{\Theta}$ of the structural parameters, the impulse-response 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

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:

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

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 LP estimates of impulse responses. [Plagborg-Møller and Wolf \(2021\)](#), [Montiel Olea and Plagborg-Møller \(2021\)](#), and [Li et al. \(2021\)](#) 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 local projections 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 VARs and the higher variance of LPs, such that shrinkage estimators—e.g. Bayesian VARs or penalized LPs ([Barnichon and Brownlees, 2019](#))—become attractive.

Our focus on long-run dynamics requires a careful consideration of this bias-

recursively by

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

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

variance trade-off. To balance these two considerations, we set the lag length of our baseline VAR to $p = 60$. This choice fulfills our desire to look at horizons well beyond the eight years traditionally associated with business-cycle frequencies while retaining as much parsimony as possible. In Section 5, we show however that our results do not depend on any specific (large) number of lags one has to select to look at the long-run.

As for inference, we take a Bayesian approach and apply priors that shrink coefficients towards zero at a rate that exponentially increases with the more distant lags, in the spirit of the “Minnesota” priors of Doan et al. (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 reduce the increase in variance stemming from the very large number of parameters involved. Moreover, by placing more shrinkage on more distant lags, the Minnesota prior can be viewed as a conservative approach to draw inference on long-run impulse responses: the data needs to speak strongly about the presence of low frequencies effects to counteract the a-priori view that these are absent. Further details on the specification of the prior are given below.

2.2 Prior specification and posterior sampling

We will use a Normal-Inverse Wishart prior over the reduced form parameters, (\mathbf{B}, Σ) . 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 et al., 2015). Denoting $\mathbf{b} = \text{vec}(\mathbf{B})$, the prior distribution is $NIW(\underline{\nu}, \underline{\mathbf{S}}, \underline{\mathbf{b}}, \underline{\mathbf{V}})$. As discussed above, we employ the “Minnesota” priors proposed by Doan et al. (1984), which shrink the VAR coefficients towards simple univariate specifications. In particular, the degrees of freedom of the prior covariance matrix are

set to $\underline{\nu} = n + 2$, with $\underline{\mathbf{S}}$ a diagonal matrix.² As for the autoregressive coefficients, the prior has the following mean and variance:

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

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

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 prior. The term $\frac{1}{\ell^2}$ implies that more distant lags are shrunk at an exponentially increasing rate towards zero. Therefore, the Minnesota prior penalizes rich large structures and favors models with shorter lags. Because of this, the choice of the tightness of the prior becomes especially important for our results about any possible long-run effect. On the one hand, if λ is large, the prior is too loose and the large number of parameters means that the long-run effects will be estimated imprecisely. On the other hand, as $\lambda \rightarrow 0$, the long-run effects are dogmatically shrunk towards zero and the data has no chance to speak about the more distant future. [Giannone et al. \(2015\)](#) propose a theoretically-grounded methodology to optimally choose the hyperparameters of the prior, based on maximization of the marginal likelihood. Based on this procedure, we select $\lambda = 0.58$ for our baseline estimates, and we will explore the results of tighter or looser choices in detail in [Appendix K](#). The conjugate nature of the prior allows us to sample from the posterior distribution in a straightforward way, using the standard

²As common, we set $\underline{\mathbf{S}}_{i,i}$ to the residual variance of a univariate AR(1) estimated on the full sample.

algorithm described in Appendix B.

2.3 Bayesian Local Projections

We compare the results of our Bayesian VAR to those based on local projections (LP). As we shall see, just like for the VAR, augmenting the local projections with a large amount of lags will be critical to appropriately recover long-run effects in our application. Therefore, Bayesian shrinkage will be needed also in the LP to reduce the variance of the estimates given the large number of parameters. To maximize comparability, we estimate equation (2) with Bayesian methods, implementing a prior on the coefficients for the lags that has the same mean and variance as in equations (3)-(4). It is important to note that while the two approaches will converge to the same results in large samples, the prior acts in a different way in the VAR and the LP. To see this, recall that one can think of both the recursive VAR identification and the lag augmented LP in terms of a two-stage approach, in which the military spending news is first regressed on p lags of the endogenous variables and itself, and then a second stage in which the impulse responses to the first-step regression residuals are calculated, either by iterating on the VAR coefficients, or by direct projection as in the LPs. The Minnesota prior in the VAR applies shrinkage to the coefficients on the lagged controls, which are then used to calculate the IRFs, therefore implicitly shrinking the latter. The Bayesian LP approach applies shrinkage to the coefficients on the control variables but does not discipline the shape of the IRFs.³

Finally, as discussed by [Miranda-Agrippino et al. \(2021\)](#), the Gaussian likelihood of model (2) is misspecified due to the presence of serial correlation in the residuals at $h > 1$. We follow these authors in interpreting it instead as the likelihood of a

³Moreover, given that the LP represents only one line of the VAR, the prior we impose is in fact an independent Normal-Inverse Wishart rather than the standard conjugate Normal-Inverse Wishart we use in the VAR. Our approach thus differs from the one proposed by [Miranda-Agrippino and Ricco \(2021\)](#), who center the priors for the LP coefficients around the IRFs produced by a low-order VAR. In our application, there is no particular reason to believe a-priori that a low order VAR is a reasonable approximation of the data, especially in the long-run.

misspecified auxiliary model. However, unlike [Miranda-Agrippino et al. \(2021\)](#), we rely on the analysis in [Montiel Olea and Plagborg-Møller \(2021\)](#), who show that lag-augmentation in LPs, as we do here, obviates the need to adjust the covariance matrix for the presence of unmodeled serial correlation. Accordingly, in our baseline estimates, we report standard Bayesian posterior density intervals.⁴

2.4 Data and identification

Our baseline data comes from [Ramey and Zubairy \(2018\)](#) and contains seven variables from 1889Q1 to 2015Q4: 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. Following [Ramey and Zubairy \(2018\)](#), we scale military news, real GDP and government spending by a measure of trend GDP, estimated as a sixth-degree polynomial for the logarithm of GDP, from 1889q1 through 2015q4, excluding 1930Q1–1946Q4. As discussed at length in the aforementioned paper, this transformation is important when computing government spending multipliers from impulse responses.

We also extend [Ramey and Zubairy \(2018\)](#)'s original dataset along several dimensions. First, we construct new series of quarterly private consumption and investment expenditures. Unpublished annual estimates of investment by the Bureau of Economic Analysis are available since 1901. Before that, we rely on the Macroeconomic History Database of [Jordà et al. \(2017\)](#), which also provide us with a measure of annual private consumption since 1890. We interpolate these series to quarterly frequency using the quarterly consumption and investment from NIPA (after 1947), [Gordon \(2007\)](#) (between 1919 and 1940) and real GDP (before 1919 and from 1941 to 1946). Second, we construct quarterly measures of Total Factor Productivity (TFP). The annual productivity series

⁴For completeness, we have verified that adjusting for residuals serial correlation produces less accurate estimates but does not overturn the significance of our results. More specifically, the estimated short-run and long-run multipliers are still significant and statistically different one from the other at the 90% level.

comes from [Bergeaud et al. \(2016\)](#), which we adjust for capital and labour utilization following [Imbs \(1999\)](#). We interpolate the annual data using the quarterly series of adjusted-TFP in [Fernald \(2012\)](#) (after 1947) and real GDP (before 1947). We also include quarterly data on US patents, provided by IFI CLAIMS Patent Services via Google Patents Public Data.

In addition, we construct new historical series of government consumption, investment and R&D spending. An annual series of government investment is available from the BEA since 1914. For the period 1890-1913, we reconstruct public investment by manually transcribing data from both the *Historical Statistics of the United States* ([Census, 1949](#)) and the annual *Statistical Abstracts* published by the census. We then interpolate this series to quarterly frequency using quarterly government spending, and back out public consumption as residual. Finally, we construct a quarterly series for public R&D expenditure relying on annual (quarterly) BEA data from 1920 (1947). For the period 1891-1919, we perform an imputation based on the fitted values from a regression of public R&D on public investment and patents. Further details are provided in [Appendix A](#).

In all cases, when temporally disaggregating a time series 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, will depend mainly on the low-frequency properties of the data, which in turn are pinned down by the properties of the annual time series. These annual series are mostly available from existing sources, and we take them at face value. The interpolation that allows us to move from annual to quarterly frequency will impact only the high-frequency properties of the data (i.e. within the year) and, as such, the specific method or the series used to interpolate is unlikely to have any effect on our estimated IRFs at longer horizons.

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. As [Plagborg-Møller and Wolf \(2021\)](#) point out, this approach yields valid impulse response estimates even if the shock of interest is non-invertible or if the instrumental variable is contaminated with measurement error that is unrelated to the shock of interest.⁵

3 Main results on military spending

In this section, we report our main results, which are based on the quarterly VAR described in the previous section using sixty lags. We begin by analyzing impulse responses and then move to the estimates of the (present value) output multipliers across forecast horizons, up to sixty quarters. Finally, we present the results of an extended VAR, where we add newly constructed time series of investment, productivity, innovation, government R&D, public investment and public consumption since 1890Q1 to shed light on the transmission mechanism of government spending.

3.1 Impulse response analysis

A simple but effective 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 point-wise 68% and 90% posterior credible sets. For ease of interpretation, the military spending news shock is normalized so as to increase government spending by 1% of GDP over the first year after the shock. The top row of [Figure 1](#) presents the impulse responses of government spending and real

⁵Furthermore, the use of a quarterly VAR(60) featuring both the instrument and the endogenous variables of interest as well as the focus on horizons up to 60 quarters ensure that our set-up meets the conditions for consistency and efficiency of the impulse response estimates provided by [Baek and Leeb \(2021\)](#).

GDP, both relative to potential output, the middle row refers to the log of the GDP deflator and the short-term nominal interest rate whereas the bottom row focuses on the government balance sheet by reporting fiscal deficit and public debt, both expressed as a share of GDP.

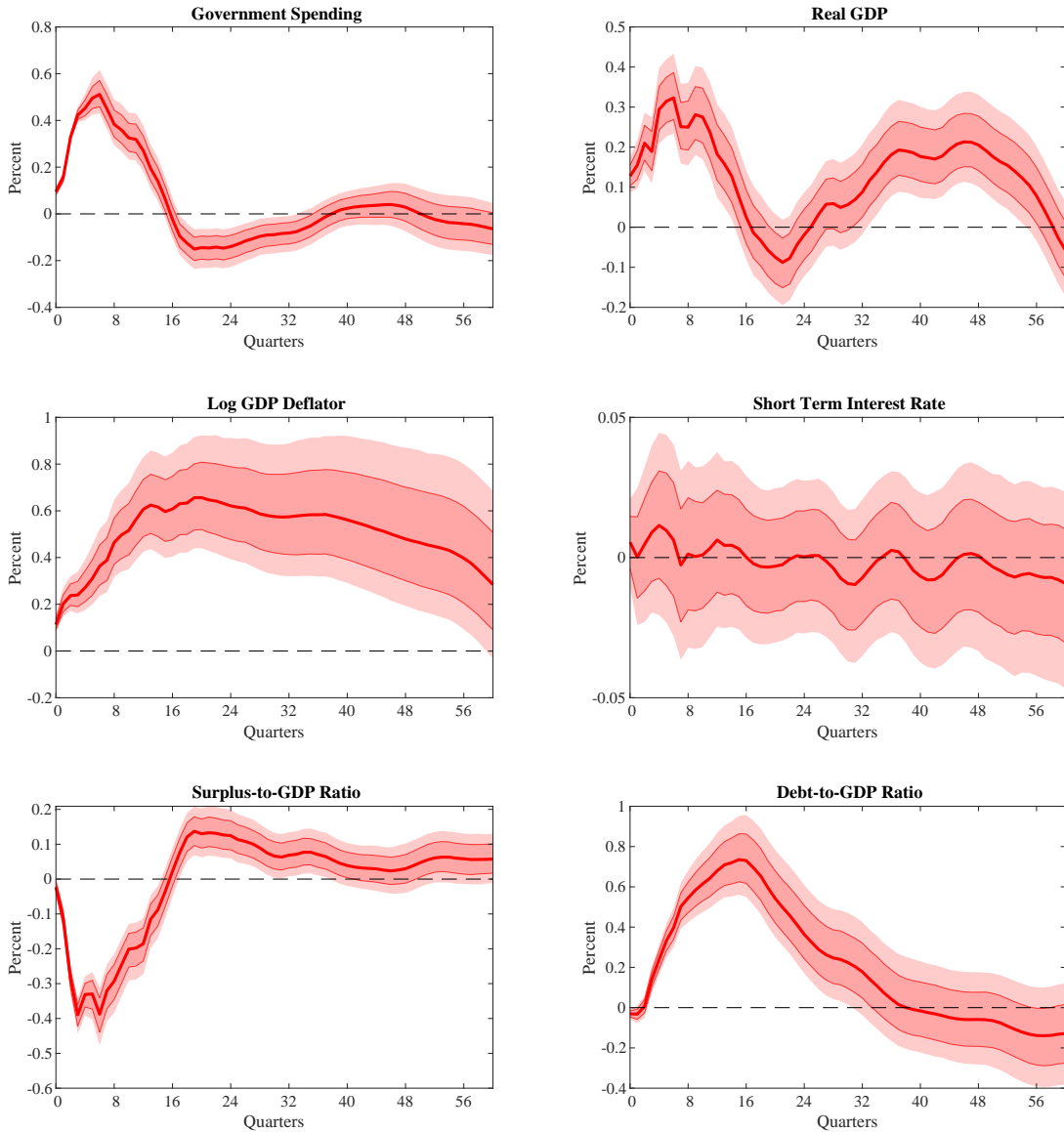
Three main findings emerge from our VAR(60). During the first four years after the shock, government spending increases sharply and then reverts to zero, triggering an equally persistent increase in GDP, a notable fiscal deterioration with government debt peaking around 0.8% of GDP, and a significant price spike above 0.6% (or 0.15% inflation per year). At frequencies between 4 and 8 years, government spending shrinks, causing a small and short-lived recession that is associated with a progressively lower price level and a fiscal surplus, which contributes to revert the path of the debt-to-GDP ratio.⁶ In the long-run, conventionally defined as frequencies beyond 8 years, the responses of government spending, surplus and debt are no longer significant but real GDP witnesses a second boom, whose peak of 0.2% is comparable to the peak in the first wave. At longer horizons, output and prices return to their historical averages, while the effects on the short-term nominal interest rate are negligible throughout. Adding a long-term interest rate, such as the yield on 10-year government bonds, as an additional variable in the VAR produces very similar findings. This is shown in Appendix D.⁷

For completeness, we report the Forecast Error Variance Decomposition (FEVD) in Appendix C. This reveals that the military spending news shock explains about 40% of the variance of government spending at business cycle frequencies, and around

⁶The sequence of fiscal surpluses in Figure 1 associated with the government spending contractions between year 4 and 10 are notably smaller than the fiscal deficits triggered by the initial government spending expansion. This suggests that the response of GDP played a significant role in reducing the debt-to-GDP ratio to pre-shock levels, consistent with the evidence in [Hall and Sargent \(2011\)](#).

⁷As noted by [Meltzer \(2004\)](#), until the Treasury-Fed accord of 1951, the Fed pegged interest rates at a low level in order to facilitate the financing of government debt during WWI, WWII and (to some extent) the Korean war. [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. The responses of the price level and the interest rates reported in Figure 1 and Figure D.1 are consistent with these historical accounts.

Figure 1: IMPULSE RESPONSES TO MILITARY NEWS SHOCK



Notes. The impulse responses are 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 and government spending are expressed in percent of potential output as defined in [Ramey and Zubairy \(2018\)](#). 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. 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, discarding explosive roots.

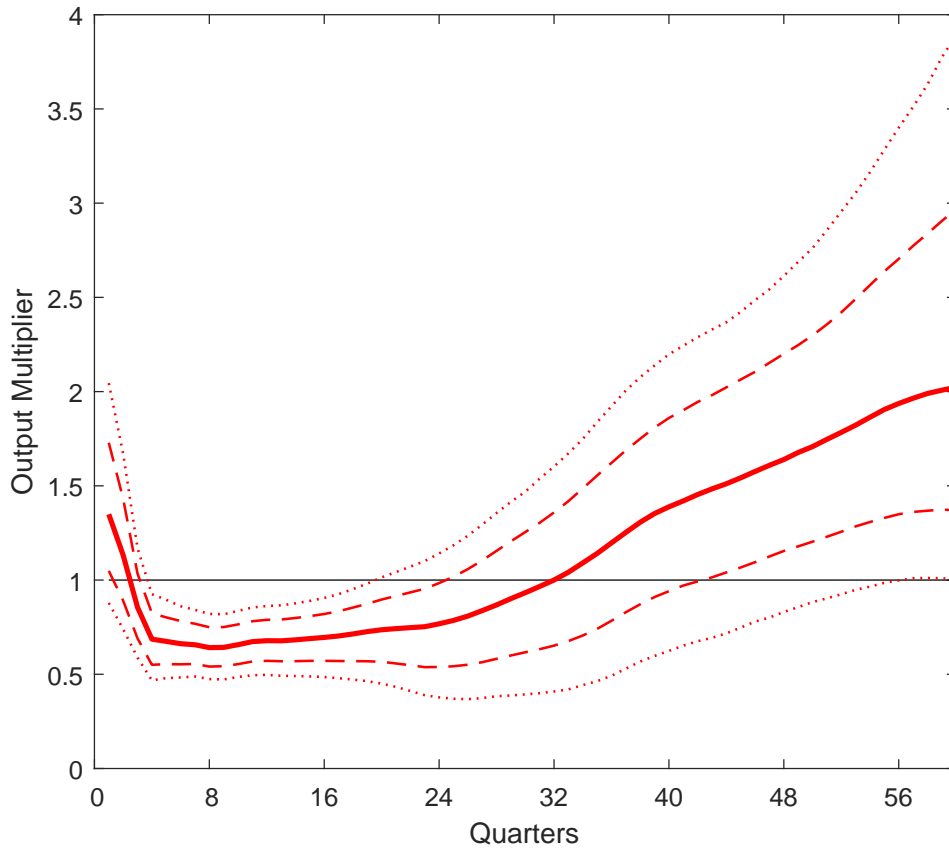
20% in the long-run. Furthermore, government spending appears to account also for a nontrivial fraction of the variance of real GDP and the price level, at about 15% and 18%, respectively.

In summary, we estimate significant long-run effects of government spending on both output and prices. Unlike the short-run dynamics where the movements in government spending are larger than the response of GDP, the lower frequency estimates suggest a large long-run multiplier as the effects on output are associated with little changes in government spending over longer forecast horizons. In the next section, we corroborate this conjecture by formally computing the multiplier across forecast horizons. Furthermore, we report the impulse responses of investment, TFP, patents and several categories of government spending to shed light on the drivers of the second wave of output effects. As for the price level, the bulk of the increase appears concentrated at business-cycle frequencies, and in particular during the first four years. Still, it takes more than a decade for prices to return to their pre-shock level.

3.2 The government spending multiplier in the short and long run

In the previous section, we have estimated a larger (smaller) output response at longer (shorter) horizons relative to the small and insignificant (large and significant) lower-frequency (higher frequency) movements in government spending. 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.

Figure 2: THE PRESENT VALUE MULTIPLIER ACROSS FORECAST 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 as in [Mountford and Uhlig \(2009\)](#). The estimates are based on a 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 and government spending are expressed in percent of potential output as defined in [Ramey and Zubairy \(2018\)](#). The broken (dotted) lines represent the central 68% (90%) HPD interval. The solid line stands for the median estimate. Results are based on 5000 posterior draws, discarding explosive roots.

In line with the empirical literature, we define the output multiplier for each horizon h as the ratio between the cumulative impulse response of real GDP to a military spending news shock 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 average nominal interest rate to discount the estimates

at each horizon between one and h quarters ahead. The findings from this exercise are reported in Figure 2, which displays the present value multiplier for each horizon between $h = 0$ (i.e. the impact multiplier) and $h = 60$ (i.e. the long-run multiplier) based on the estimated parameters of the VAR(60).

The figure reveals that the government spending multiplier is, on impact, about 1.35, with most of the distribution mass above one. After the first four quarters, however, the multiplier decreases to values significantly below one, around 0.7, consistent with the evidence in Hall (2009), Barro and Redlick (2011) and Ramey and Zubairy (2018). These estimates are relatively stable over the following five years before growing with the forecast horizon. The posterior median of the multiplier takes values above one at frequencies beyond thirty-two quarters and peaks at the significantly larger value of 2 in the forecasts fifteen years ahead, despite the deterioration of accuracy due to the longer horizons.

In summary, our results suggest two main conclusions. 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 literature, between 0.6 and 1.5, thereby offering a reconciliation of apparently conflicting findings in earlier studies. Second, while the multipliers at business-cycle frequencies are statistically below one, the long-run multipliers are much larger and eventually exceed one significantly.

3.3 Inspecting the mechanism

The findings in the previous section uncover a significant second wave of output response at longer horizons to a temporary government spending shock. To shed light on the transmission mechanism, in this section we look at the effects of the military spending news shock on investment, productivity and innovation, as measured by the number of granted patents per quarter. Furthermore, we add public R&D, government investment and government consumption to our baseline VAR(60) to explore the role

of different public spending categories. This richer empirical specification comprises thirteen variables.⁸

TFP and patents enter the VAR in log-levels. This introduces non-stationarity relative to our baseline model. Moreover, while most components of government spending appear stationary once expressed as a ratio to potential GDP, public R&D over potential grows exponentially during the sample and thus displays non-stationarity even after taking logarithms. For this reason, we generalize the prior in Section 2.2 to the “Dummy Initial Observation” extension (Sims, 1993; Sims and Zha, 1998), which is designed to handle a mix of stationary and (possibly cointegrated) non-stationary variables in the VAR.⁹

In Figure 3, we report the posterior credible sets of the responses to a government spending shock based on the extended VAR(60). The top row focuses on investment and productivity, the middle row refers to innovation and government R&D while the bottom row records public investment in equipment, structures, and software (which for lack of a better term we label ‘public non-R&D investment’) and public consumption. In the top right panel, we report in red (grey) the response of productivity from the extended VAR(60) in which the series of TFP is (un)adjusted for capital and labour utilization. Given that variations in utilization rates are concentrated at higher frequencies, we would expect this adjustment to have little impact on the estimated long-run effects.

Starting from the top row, a number of interesting results emerge. The short-run response of investment (and consumption, shown in Appendix Figure D.2) is negative and significant. This helps rationalize why the response of GDP is smaller than the response of government spending at business-cycle frequencies, thereby generating

⁸The curse of dimensionality prevents us from extending further the variable set of the VAR(60) estimated in this section. In Appendix D, however, we present the estimates of our baseline VAR(60) augmented with private consumption and investment, in a parallel to the augmented VAR used in Section 3.1 to look at the long-term rate.

⁹The parameter $\theta = 0.05$ which controls the tightness of this prior is also selected by maximizing the marginal likelihood following Giannone et al. (2015). As discussed in Section 6, our baseline estimates of Figure 1 are robust to using the “Dummy Initial Observation” prior of (Sims, 1993; Sims and Zha, 1998).

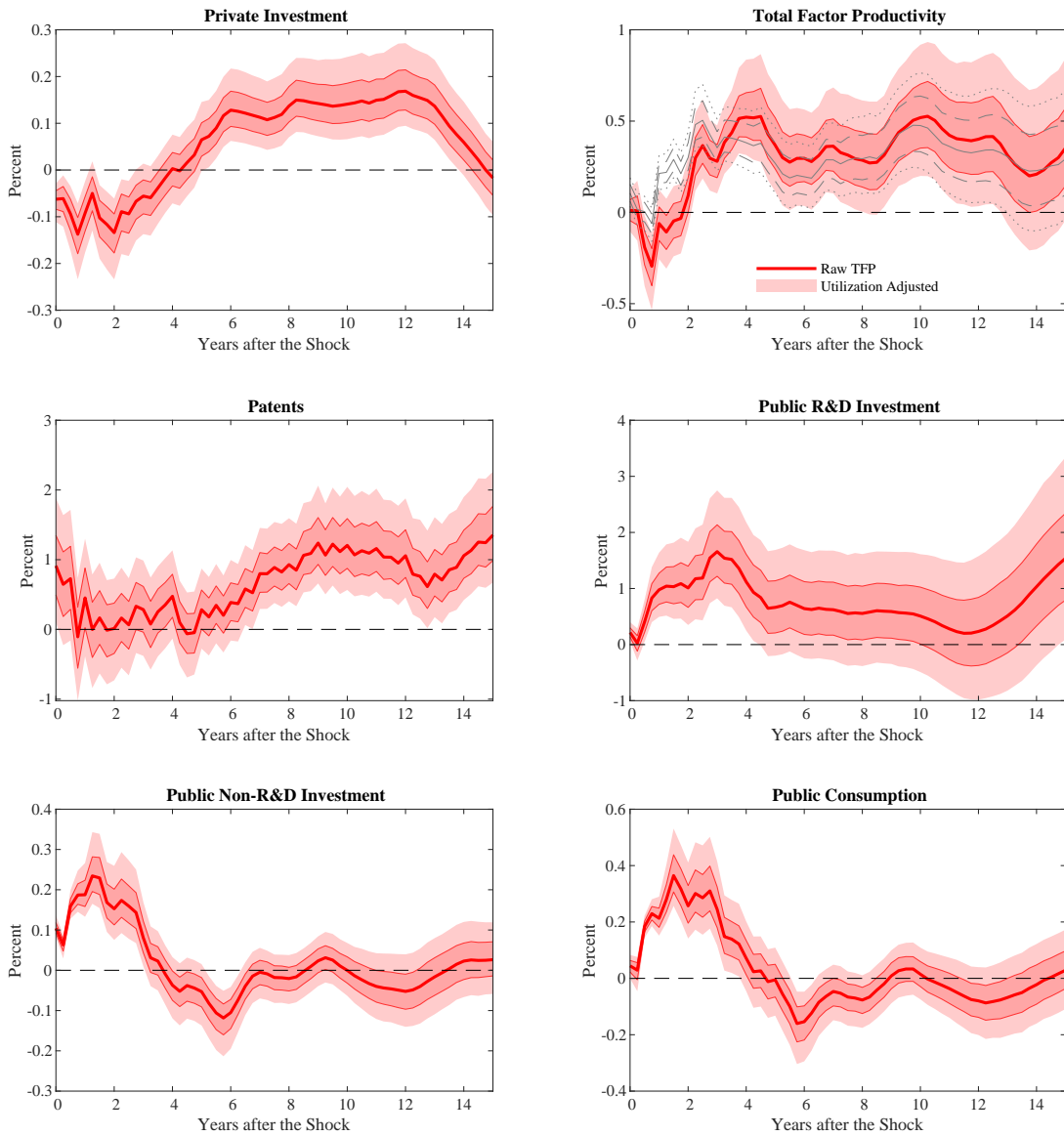
a short-run multiplier below one. However, at the end of the fourth year, when government spending shrinks relative to its historical average, the investment response turns positive and significant for a sustained period of time. Accordingly, the output multiplier exceeds one in the long-run. In other words, our estimates suggest that government spending crowds private investment out at shorter horizons but crowds it in at longer horizons.

Moving to the top right panel reveals that, after a negligible impact during the first two years, the response of TFP becomes significant and long-lasting, returning to a statistical zero around quarter 52. The utilization adjustment bears no material implication for our long-run estimates, as exemplified by the finding that the red and grey bands largely overlap at lower frequencies. Similarly, the left panel of the middle row of Figure 3 shows that the response of patents is economically and statistically small over the first five years but, after that, the effects of military spending on innovation are strongly significant and long-lasting, consistent with the evidence in [Diebolt and Pellier \(2022\)](#).

In the rest of Figure 3, we look at different components of government spending. The right column of the middle row makes it clear that military spending triggers a strong and sustained increase in public R&D, which is particularly significant at the beginning and at the end of the forecast horizon. In contrast, the bottom row reveals that the responses of public (non-R&D) investment and public consumption are much shorter-lived: the initial humps are reversed by year 5, briefly turn negative, and, thereafter, the changes in either category are no longer significant. Interestingly, Figure 3 points to a strong correlation between public investment and public consumption, suggesting that military spending drives a strong co-movement between all three categories, especially at shorter horizons. We will come back to these correlations in the next section.

The timing of the estimated dynamic effects on investment, TFP and patents

Figure 3: IMPULSE RESPONSES TO MILITARY NEWS SHOCK



Notes. The impulse responses are based on an estimated VAR with sixty lags of military spending news, real GDP, GDP deflator, short-term interest rate, fiscal deficit to GDP ratio, government debt to GDP ratio, real private investment, utilization-adjusted TFP, patents, public consumption, public non-RD investment and public RD spending. Military spending news is ordered first in the Cholesky factorization. Output, private investment, and government spending variables are expressed in percent of potential output as defined in [Ramey and Zubairy \(2018\)](#). 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% (90%) HPD band. The darker solid line stands for the median estimates. Results are based on 5000 posterior draws, discarding explosive roots.

suggests a leading role for the corporate sector in amplifying the effects of government spending on GDP in the medium- and long-term. On the other hand, while both public investment and public consumption increase significantly over the first four years, only the rise in government R&D is highly persistent. In the next section, we will ask whether an exogenous increase in public R&D can have long-lasting effects on productivity and innovation *on its own* (i.e. without being triggered by a surge in military spending). We will contrast the estimates of this exercise with the dynamic effects of a shock to either public (non-R&D) investment or public consumption.

4 Is defense special? Evidence from other spending categories

In the previous section, we have shown that (i) military spending is associated with a persistent increase in public R&D; (ii) military spending boosts productivity and innovation in the medium term; (iii) the increase in productivity and innovation leads the second wave of output effects. These findings suggest an important role for public R&D in shaping the long-run effects of defense spending. At the same time, our analysis begs the important question of whether there is anything ‘special’ about military spending that triggers these chain reactions, or perhaps an exogenous increase in public R&D (or possibly any other government spending category) could achieve similar outcomes, without being necessarily associated with an expansion in the defense budget.

In this section, we design a simple strategy that, in the spirit of [Perotti \(2004\)](#) and [Ilzetzki et al. \(2013\)](#), tackles this issue. Our starting point are the impulse responses of the augmented VAR of the previous section, which highlight a strong contemporaneous co-movement between public R&D, public investment and public consumption, following a military spending news shock. The empirical strategy that we propose attributes, in turn, the entirety of the contemporaneous correlations

between public R&D, public investment and public consumption to the effects of an exogenous increase in: (i) public R&D, (ii) public (non-R&D) investment and (iii) public consumption, respectively. In practice, we construct a “shock” to each public spending category using a Cholesky factorization in which the public spending category of interest is ordered first in the VAR(60). For instance, this corresponds to assuming that a public R&D shock is the only shock that can influence public R&D spending contemporaneously, which is: public investment and public consumption (as well as GDP or any other variables, following [Blanchard and Perotti, 2002](#)) cannot affect public R&D spending *within the quarter*. In contrast, the public R&D shock is allowed to affect public investment and public consumption *on impact*. Accordingly, this ‘scenario’ implicitly assigns the whole contemporaneous variation in public R&D to a public R&D innovation, and the entirety of the contemporaneous correlation among public spending categories to the effect of government R&D spending on public investment and public consumption.

In the same vein, we construct a public (non-R&D) investment shock and a public consumption shock by using a Cholesky factorization but, this time, we order—in turn— public (non-R&D) investment and public consumption first. This ‘scenario’ attributes the impact variation in public investment and in public consumption to a public investment shock and to a public consumption shock, respectively, and the correlations among spending categories to the effects of either public investment or public consumption.

If the three components of government spending were perfectly correlated, the impulse responses resulting from these three ‘scenarios’ would look identical, and we would not be able to make any conclusion about which component accounts for most of the long-run effects on GDP. To the extent that these public spending categories are not perfectly correlated, however, differences in the impulses responses across the three spending category shocks could be informative about the transmission mechanism. If,

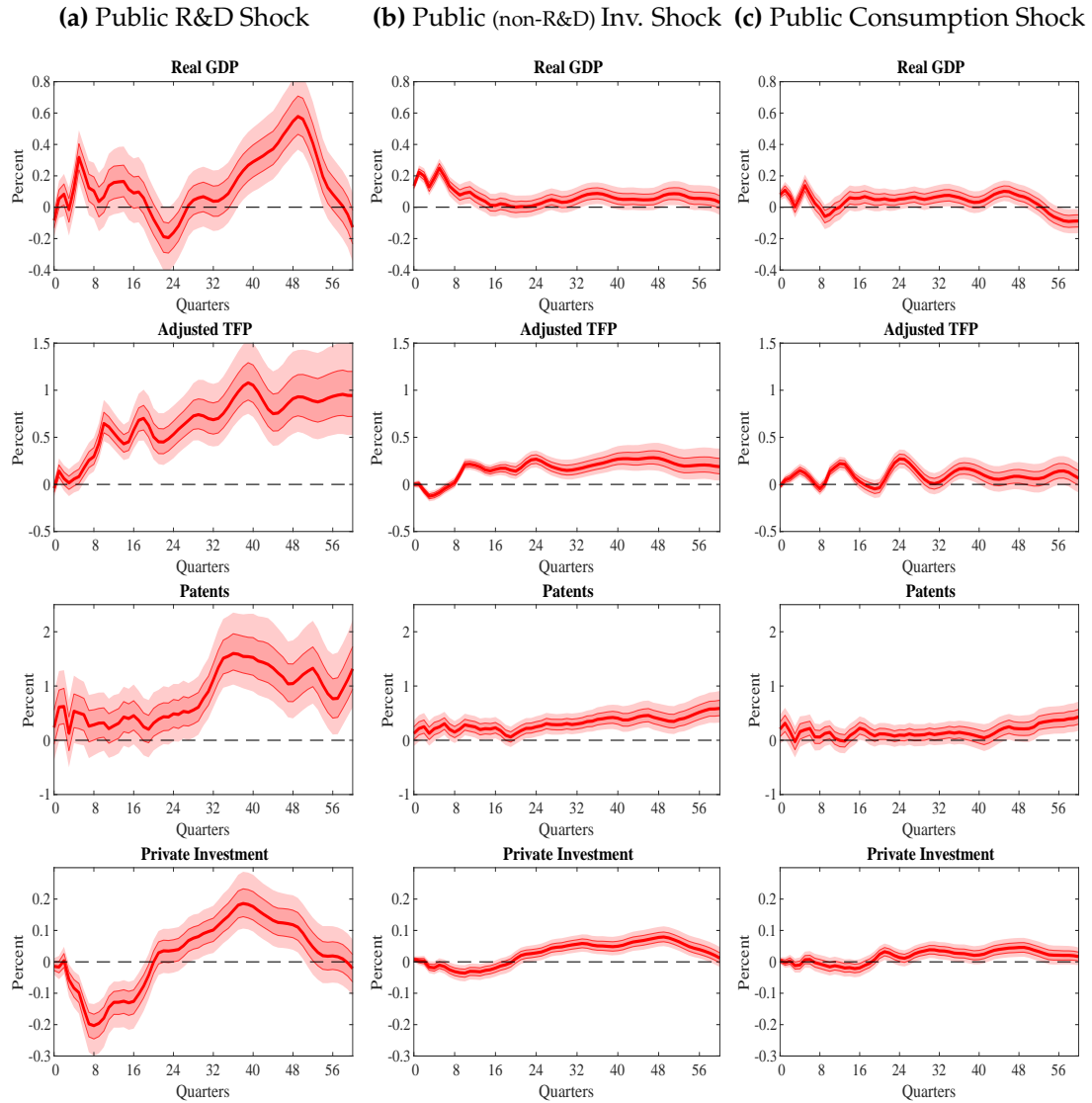
for instance, we found that: (i) the responses of productivity, innovation and output to a public R&D shock are similar to the long-run effects of the military spending news shock, but (ii) public investment and public consumption shocks triggered only modest dynamic effects on productivity, innovation and output at longer horizons, then we could conclude that public R&D is a key transmission channel for the effects estimated in Figure 3.

The findings from these three exercises are reported in Figure 4 and Appendix E. Each column refers to the effects of a “shock” to public R&D (left), public non-R&D investment (middle) and public consumption (right). The rows report GDP (top), TFP and patents (center), private investment (bottom) and prices (Figure E.1). We report median estimates together with 68% and 90% posterior credible sets. For comparability, the size of the shocks in each column is normalized such that it increases the sum of public (non-R&D) investment and public consumption by 1% of GDP over the first year after the shock.

Three main conclusions can be tentatively drawn from Figure 4. First, an exogenous increase in public R&D in the left column has large and significant effects on GDP, productivity, innovation and investment at longer horizons, but the response of the top three variables (investment) is relatively smaller (negative) at higher frequencies. Second, a shock to public consumption in the right column has no long-run effects and generates only a small short-run stimulus. Third, an innovation to public (non-R&D) investment triggers small but significant responses of GDP, TFP, patents and private investment at longer horizons and cause a short-run expansion in output that is significantly larger than the short-run effects of government consumption, consistent with the empirical evidence in Auerbach and Gorodnichenko (2012) and the theoretical model in Barro (1990).

In summary, the impulse responses to a public R&D spending shock in Figure 4 very much resemble the dynamic effects of a military spending news shock in Figure

Figure 4: THE EFFECTS OF SHOCKS TO DIFFERENT PUBLIC SPENDING COMPONENTS



Note: Solid lines are median posterior responses. The lighter (darker) shadow area represents the 95th (68th) HPD interval. In all three columns, the reduced-form VAR includes sixty lags of real GDP, GDP deflator, treasury bill rate, deficit to GDP ratio, debt to GDP ratio, government consumption, government non-R&D investment and government R&D investment. In each of the columns the first variable is, respectively, public R&D investment, public non-RD investment, or public consumption expenditure, and the shock to each component is identified by means of a Cholesky decomposition. For comparability across columns, all shocks are normalized such that the sum of public consumption and public non-R&D investment is as large as 1% of GDP over the first year after the shock.

1, especially in the medium- and long-term. But the changes triggered by a public investment shock or a public consumption shock do not look at all like the impulse responses to a military spending shock, except for the first two-years. We interpret these findings as suggestive evidence that the effects of government spending on GDP, productivity and innovation at longer horizons operate mostly through their impact on public R&D, whereas the short-run effects on GDP are mostly accounted by public investment. Overall, public consumption appears to play a far smaller role, if any.

5 On the importance of going (very) long

In the previous part of the paper, we have documented large and significant effects of government spending on output and prices at low frequencies, based on a quarterly VAR with sixty lags. In this section, we argue that including a large number of lags is important to uncover the long-run effects of fiscal policy. To do so, we will compare the estimates of the long-run multiplier estimate (i.e. $h = 60$ quarters ahead) obtained by VARs and LPs of different lag length. As we will see, the differences in results will be primarily driven by the number of lags used in each method, rather than by whether one uses VARs or LPs.

In addition, we will show the extent to which our results reflect the fact that the military spending news variable can be predicted at longer horizons. Finally, we will rely on simulation analyses to show that a VAR with four lags, a widely-used specification in empirical macroeconomics, is not able to successfully capture the ‘true’ (long-run) dynamic effects of government spending when the data generating process (DGP) contains sixty lags. In contrast, we find that a VAR(60) correctly recovers the ‘true’ IRFs —both at shorter and longer horizons— when the DGP is a VAR(4).

5.1 Estimates based on (much) shorter lag lengths

In Table 1, we report the present value long-run multiplier, \mathcal{M} , based on two sets of VAR (left panel) and LP (right panel) specifications that differ only by their lag length p . For ease of comparability, we set the shrinkage parameter, λ , in each row equal to 0.58, which is the value that maximizes the marginal likelihood in the VAR(60). Within each panel, the columns display the 5th, 50th and 95th percentiles of the posterior distribution of the long-run \mathcal{M} as well as the share of draws for which the long-run multiplier is above one.

The left four columns of the last row of Table 1 report the findings from our baseline VAR(60). The long-run multiplier is 2.1, with a credible set of 1 to 3.9, and corresponds to the estimates at the end point of Figure 2. The four columns of the last row on the right of Table 1 are based on local projections that control for sixty lags of all variables. A robust result that emerges from the last row of the table is that the likelihood of a long-run multiplier above one is very significant, independently of the method used. The share of posterior draws for which $\mathcal{M} > 1$ is 96% and 100% for the VAR(60) and LP(60) respectively.

To appreciate the impact of lag length selection, in all other rows of Table 1, we report the present value long-run multiplier using VAR(p) and LP(p) specifications in which the number of lags, p , becomes progressively smaller moving up the table. The first row of Table 1 computes the long-run multiplier, \mathcal{M} , using a VAR(4) and a LP(4), respectively, as typically done in the empirical macro literature using quarterly data. According to these specifications, the posterior median is much smaller, between 0.7 and 0.9, with a tighter credible set of [0.3, 1.2], probably due to the far smaller number of parameters. More importantly, the probabilities that the long-run $\mathcal{M} > 1$ are now only 12% and 37%.

The remaining rows of Table 1 show the long-run multiplier for different lag length

Table 1: THE PRESENT VALUE LONG-RUN MULTIPLIER, \mathcal{M}

No. lags (p)	Vector AutoRegressions (VAR)				Local Projections (LP)			
	5^{th}_{pct}	50^{th}_{pct}	95^{th}_{pct}	$\mathcal{M} > 1$	5^{th}_{pct}	50^{th}_{pct}	95^{th}_{pct}	$\mathcal{M} > 1$
4	0.28	0.67	1.19	12%	0.72	0.95	1.21	37%
10	0.52	1.00	1.74	49%	0.85	1.13	1.48	77%
20	0.45	1.04	1.90	54%	0.68	0.93	1.23	34%
30	0.35	0.95	1.73	44%	1.13	1.58	2.29	99%
40	0.53	1.77	4.21	83%	2.73	3.74	5.65	100%
50	0.94	2.30	5.14	94%	2.11	2.96	4.46	100%
60	1.03	2.08	3.90	96%	1.73	2.26	3.01	100%

Notes: The long-run multiplier is computed over a forecast horizon of 60 quarters. The Vector AutoRegression (VAR) and Local Projection (LP) specifications use seven variables: 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 of the VAR. The definitions of the variables and the present value multiplier, \mathcal{M} , follow [Ramey and Zubairy \(2018\)](#) and [Mountford and Uhlig \(2009\)](#). Results are based on 5000 posterior draws, discarding explosive roots. Each row refers to the estimates of a different specification of either the VAR(p) or the LP using p lags of all variables. The number of lags, p , for each specification is reported in the first column. The columns 5^{th}_{pct} , 50^{th}_{pct} and 95^{th}_{pct} present the 5^{th} , 50^{th} and 95^{th} percentiles of the posterior distribution of the present value multiplier \mathcal{M} . The columns $\mathcal{M} > 1$ report the share of posterior draws for which the ratio of the cumulated response of GDP and the cumulated response of government spending is larger than one.

choices that run in between the two extremes of 4 and 60. A few results emerge from this exercise. First, the posterior median of long-run \mathcal{M} tends to increase with the lag length, reaching values around 1 with ten lags and exceeding 1.7 (1.5) using forty (thirty) lags in the VAR (LP). Second, the accuracy of the point estimates progressively deteriorate with the sharp increase in the number of parameters associated with the richer specifications. Third, and notwithstanding the larger uncertainty just discussed, the likelihood of a long-run multiplier above one is a positive function of the lag length, with the largest increases in the left panel recorded for the VAR(10) and VAR(40): adding only six lags brings the probability of $\mathcal{M} > 1$ in the long-run from 12% with the VAR(4) to 49% with the VAR(10); but it takes another thirty lags to further reach the 85% of the VAR(40).

A very similar picture emerges from the right panel Table 1, which collects results from the LP models. The posterior median of the long-run multiplier tends to be a positive function of the number of lags, p , with the largest rises in the share of draws for which $\mathcal{M} > 1$ recorded when moving from the LP(4) to the LP(10) specification and

then from the LP(10) to the LP(30). Relative to the VAR estimates, the LP specifications in the right panel of Table 1 seem to deliver slightly higher and more accurately estimated multipliers.

For completeness, in Appendix Table H.1, we report the equivalent of Table 1 but for three popular forecast horizons used in the literature on the fiscal multiplier: one quarter (Panel A), two years (Panel B) and four years (Panel C). Interestingly, at these much shorter horizons, the various specifications perform very similarly, independently of the number of lags or of using VARs versus LPs. The median impact multipliers range from values of 1.2 to 1.4, with most of the posterior distribution mass above one. In contrast, the central estimates of the present value multipliers at two- and four-year horizons are always significantly below one, spanning the interval between 0.6 and 0.8.

5.2 Low-frequency predictability

At this stage, we find it useful to look at the full set of impulse responses based on the VAR(4) and the LP(4), given that these specifications are widely used in studies on quarterly data. The results are shown in Appendices F and G. The robust finding that emerges from this comparison is that both four-lag specifications miss the large and significant long-run effects of government spending on output and prices that we have estimated in the previous sections on the basis of models with a far more generous lag structure.

One way to understand the discrepancy between low- and high-order VARs or LPs is to realize that the measurement of the shock that is projected onto the endogenous variables in both methods differs depending on the number of lags employed as controls. Both in the case of the VAR, where the military news series is ordered first within a Cholesky factorization, and in the lag-augmented LP, the identified shock corresponds to the residual of a regression of military spending news on p lags of itself and all other variables. If lagged endogenous variables help predict the military

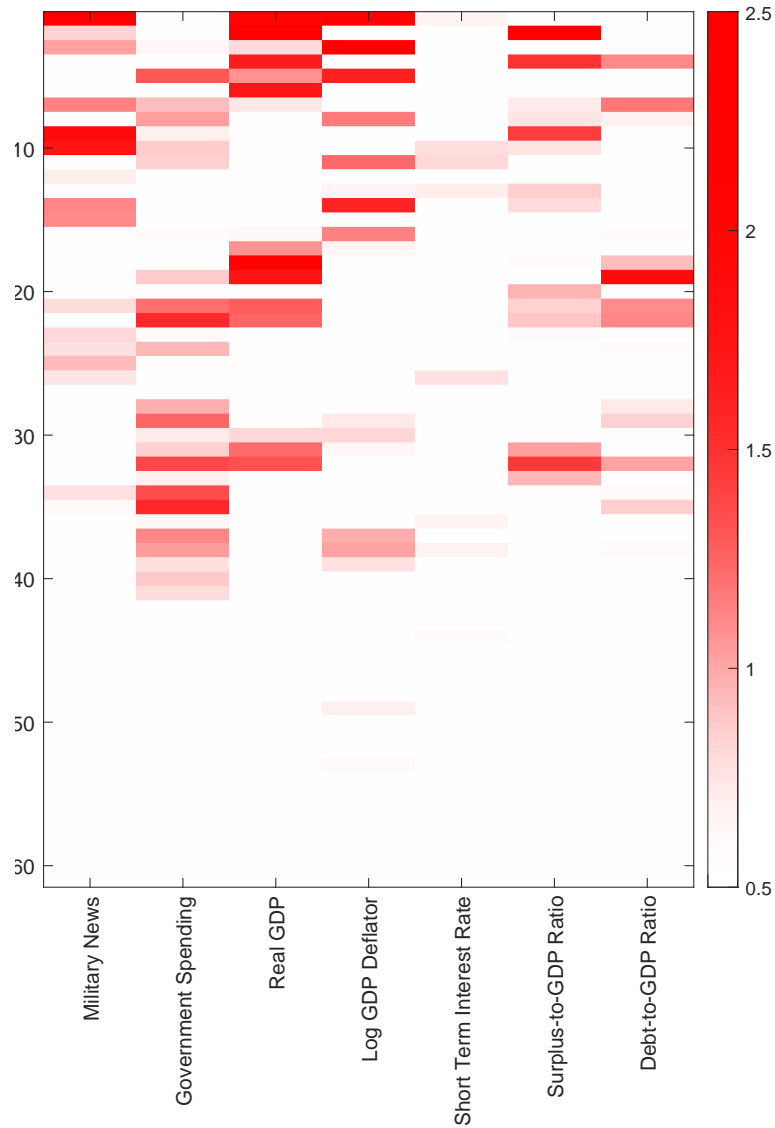
spending news (i.e. if they “cause” the military spending news in the sense of [Granger, 1969](#)), then controlling for those lags will affect the results in a classical omitted variable problem.

In [Figure 5](#), we report the outcome of regressing military spending news on 60 lags of itself and all other variables, using the same prior shrinkage described in [Section 2.2](#). The residual of this regression corresponds to the identified shock used by both the VAR(60) and the LP(60). In the figure, row i and column j jointly identify the i -th lag of the j -th variable, and the shade of the color grows with the ratio between the absolute value of the associated regression coefficient posterior mean and its posterior standard deviation. A darker shade indicates greater significance in predicting future military spending news.

Two main results can be taken away from [Figure 5](#). First, the military spending news are highly predictable, especially using lags of the military spending news itself, and every other variable except perhaps the short-term interest rate. Second, the most systematic pattern that emerges from [Figure 5](#) refers to the high significance of the estimated coefficients on lags up to about 40 quarters. Beyond that horizon, however, all lagged endogenous variables appear to lose their ability to predict future military spending news.

Two comments are worth noting. First, the findings in [Figure 5](#) suggest that a generous lag length selection is not only desirable (as shown in the previous sections) but it may be, in fact, also necessary to isolate exogenous movements in military spending news, and therefore in government spending. Second, the omission of longer lags may also be responsible (over and above any possible truncation bias) for the inability of conventional specifications such as the VAR(4) and LP(4) to identify long-run dynamics. It should be noted, however, that the omitted variable bias hinted by [Figure 5](#) seems to have little impact on the estimated effects of government spending on output at shorter horizons based on specifications with very few lags, as revealed

Figure 5: SIGNIFICANCE OF LAGS



Notes. Military spending news is projected on sixty lags of itself, government spending, GDP, GDP deflator, treasury bills, deficit to GDP ratio and debt to GDP ratio. Columns refer to the regressors and rows refers to their lags. Darker shades of red indicate higher predicting power as measured by a higher value of the ratio between (the absolute value of) the posterior mean of the estimated coefficient and its posterior standard deviation.

by a comparison between Figure 2 and Table H.1.¹⁰

¹⁰The notion of low-frequency predictability used in this paper is related to but it is distinct from the notion of low-frequency covariability used by Müller and Watson (2018). In that paper, the authors are interested in drawing inference on the *contemporaneous* relationship between the low-frequency components of two time-series, whereas

5.3 Forecast encompassing

A main finding of this paper is that empirical time series models with many lags detect significant long-run effects of government spending on output, but empirical models using far lesser lags miss these low-frequency dynamics. A legitimate question is therefore which lag length should be preferred on statistical grounds. In this section, we address this issue by developing a forecast encompassing strategy in the spirit of [Chong and Hendry \(1986\)](#).

In the first step, we use the estimated parameters of the VAR(4) to simulate data in which there are no significant long-run effects. We then estimate the VAR(60) on the simulated data to assess whether the VAR(60) incorrectly detects spurious long-run effects that are actually absent in the VAR(4) data generating process. In the second step, we do the reverse and ask whether the VAR(4) can correctly detect the long-run effects that are present when we simulate data using the estimated parameters of the VAR(60). If the VAR(60) correctly identifies no long-run output responses when the data are generated using the estimated VAR(4) (i.e. when there are no long-run effects) *and* the VAR(4) incorrectly identifies no long-run output responses when the data are actually generated using the estimated VAR(60) (i.e. when there are long-run effects), then we conclude that the VAR(60) encompasses the VAR(4) and therefore the former should be preferred.

In [Appendix I](#), we show that while the estimates of the VAR(60) on the data simulated using the VAR(4) estimates on actual data are able to replicate the insignificant long-run effects of the VAR(4) data generating process, the estimates of the VAR(4) miss entirely the long-run effects in the data simulated using the VAR(60) estimates on actual data. In addition, the VAR(4) tends to over-estimate the short-

here we are interested in the cross-frequency correlations among a set of variables, at potentially very long leads and lags. For instance, the estimated impulse responses presented in [Figure 1](#) show that movements in government spending at short-horizons can covary significantly with movements in GDP at long-horizons, even though the long-run response of government spending is virtually uncorrelated with the long-run response of GDP.

run effects on output embedded in the VAR(60) data generating process whereas the credible sets of the VAR(60) include the ‘true’ IRFs of the VAR(4) simulated data, even at shorter horizons. In summary, a further reason to prefer a generous lag length selection in VAR(p) is that specifications with a larger p are likely to encompass the long-run forecasts of specifications with a far smaller p , both at long and short horizons.

6 Sensitivity analysis

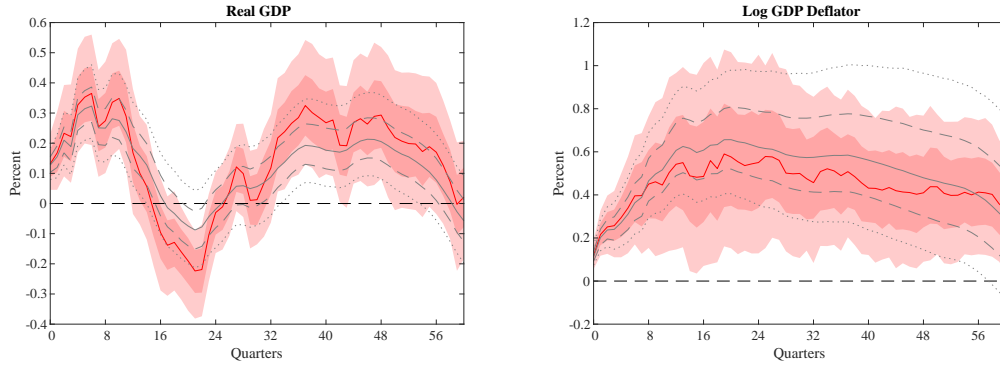
The empirical findings in the previous sections are robust to a number of sensitivity checks, including the use of local projections rather than VARs to construct the impulse response functions, varying the specific starting or ending point of the long historical sample, excluding specific war episodes, imposing the identifying restrictions in [Blanchard and Perotti \(2002\)](#) on total government spending rather than using military spending news, choosing different priors and making different assumptions about the GDP trend. In this section, we summarize these robustness exercises.

Local projections. In a recent contribution, [Plagborg-Møller and Wolf \(2021\)](#) demonstrate that, in large samples, VAR and Local Projections (LP) estimate the same impulse response functions whenever the number of lags selected in the VAR is as large as the number of quarters ahead in the forecast horizon. Furthermore, [Montiel Olea and Plagborg-Møller \(2021\)](#) show that to obtain robust inference over longer horizons, local projections should be augmented with a sufficient number of lags of all relevant control variables.

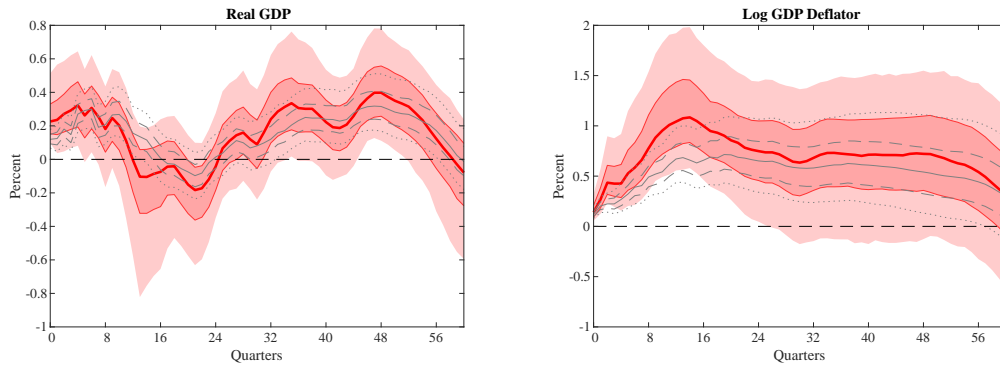
The discussion above suggests that a specification with sixty lags would be an ideal local-projection-counterpart of the VAR(60) that we have estimated in Section 3. Indeed, the findings in Table 1 confirms that the long-run multiplier estimated by the LP(60) specification (and indeed any other lag length selection p) is very similar to the estimates based on the VAR with sixty (p) lags. As discussed above, to maximize

Figure 6: IMPULSE RESPONSE FUNCTIONS UNDER ALTERNATIVE SPECIFICATIONS

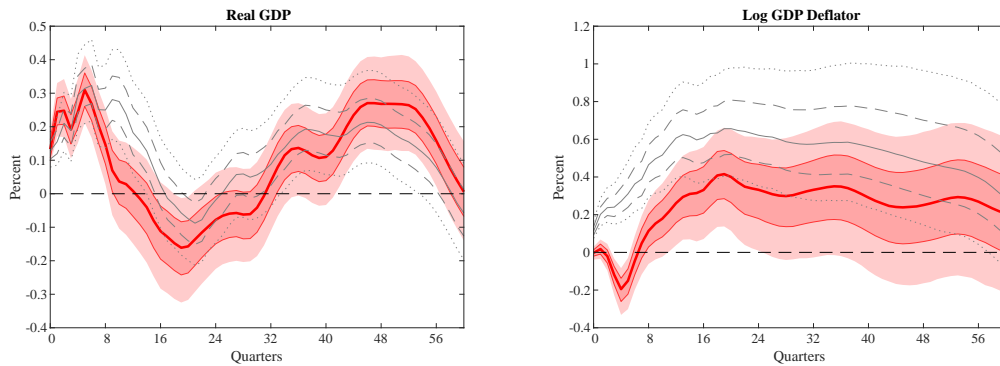
(a) Bayesian Local Projections



(b) Excluding World Wars



(c) Blanchard-Perotti's Identification



Note: The solid lines represent the median posterior response. The darker shadow area represents the 68th posterior credible intervals, while the lighter shadow are represents the 95th posterior credible intervals. 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, with the exception of the last row which excludes military spending news.

comparability across the two methods, we set the priors for the LP parameters to the same prior distributions described in equation (4) for the VARs, and choose a shrinkage of $\lambda = 0.58$.

The impulse responses estimated with local projections are reported in the top row of Figure 6 and confirms, by and large, the estimates based on the VAR(60) in Figure 1, whose bands are reported in grey for the readers' convenience. The main finding is that the shape and significance of the credible sets for both output (in the left column) and prices (on the right) based on the LP(60) are very similar to those from the VAR(60). In particular, it is still the case that: (i) the long-run impact on GDP is statistically significant also using local projections (with a generous lag length selection), and (ii) the sharp and significant price increase is short-lived but take more than ten years to fully revert. The full set of estimates associated with the LP(60) specification, including both the impulse response analysis for all variables and the present value multiplier reported as a function of the forecast horizon, are presented in Appendix J.

Excluding World Wars. As argued by [Friedman \(1952\)](#) and [Ramey and Zubairy \(2018\)](#), the use of military spending (news) and wars for the purpose of identifying the effects of government spending 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 feedbacks running from GDP to government spending. Second, these public spending swings tend to be large in historical perspective, thereby fulfilling the econometrician's desire to observe sufficient variation in the leading variable. On the other hand, using wars as source of exogenous variation poses some external validity challenges on whether a specific episode drives the empirical findings and on whether the identified effects generalize to other components of public expenditure (as shown in the previous section for the case of government R&D spending) as opposed to referring only to

military spending. Moreover, the timing of the main three large wars in the sample (the two world wars and the Korean war) is such that there are between fifteen and twenty years in between the end of each major conflict and the start of the next, raising the possibility that we are detecting a spurious cycle of war, military spending, and GDP growth.

To ameliorate some of these concerns, in the middle row of Figure 6, we run our baseline VAR(60) censoring to zero the observations of the military spending news for WWI and WWII. In Appendix L, we exclude instead the Korean war. In neither of these exercises, the exclusion of one (or any pair) of these war-induced military spending news overturns our main conclusions: (i) the long-run effects on GDP are large and significant, (ii) prices increase sharply over the first twelve quarters after the shock and then revert slowly to their historical average after fifteen years. Excluding all three war episodes at once, in contrast, produces small and insignificant impulse responses for government spending, output and prices, consistent with the notion that wars provide a significant source of variation to identify the effects of defense spending programmes.¹¹ In other words, each and every one of these unprecedentedly large events seems sufficient to elicit significant short-run and long-run effects of military spending on output and prices, though none of them is actually necessary.

Blanchard-Perotti's identification. Another popular strategy to isolate exogenous variation in fiscal policy has been proposed in the influential study by [Blanchard and Perotti \(2002\)](#). The authors identify government spending shocks by assuming that while output can respond contemporaneously to movements in public expenditure, governments take at least one quarter to adjust their level of spending in response to movements in GDP. As discussed by [Ramey \(2011b\)](#) and [Ramey and Zubairy \(2018\)](#),

¹¹In the next sensitivity analysis, based on the identification in [Blanchard and Perotti \(2002\)](#), we estimate similarly large and significant long-run effects on output and prices when using total government spending. Furthermore, in section 4, we have provided evidence of significant long-run effects on output following an increase in public R&D spending that is not necessarily triggered by an exogenous expansion in the defense budget.

this corresponds to a Cholesky decomposition with government spending ordered before GDP in the VAR. We follow this specification and use the same priors employed in the rest of the paper.¹²

The estimates based on Blanchard-Perotti's identification are displayed in the bottom row of Figure 6, and corroborate the findings from the other rows as well as the baseline results.¹³ After a government spending shock, the first rise in GDP lasts for only about eight quarters and then, at frequencies beyond 32 quarters, it emerges a second wave that seems larger, significant and more persistent than the first wave. On the other hand, the significant price response is concentrated between years 2 and 6, though relative to the military spending news identification, the impact seems delayed and the peak smaller (0.4% vs 0.6%). Overall, this alternative identification paints a similar picture relative to rest of our analysis. The finding that also the estimates using the Blanchard-Perotti identification, which does not rely on military spending, lead to a low-frequency boom in output suggests that the dynamically rich responses to government spending shocks uncovered by our analysis represent a genuine pattern, which extends beyond military spending and the military spending news identification.

Prior Tightness. All the results so far have been based on the prior tightness selection strategy described in Giannone et al. (2015), who propose to treat λ in equation (4) as a hyperparameter to be estimated in a hierarchical manner. The authors recommend setting λ to the value that minimizes the marginal likelihood of the model. In our context, this choice is attractive for at least two reasons. First, since the marginal likelihood is closely related to the one-step ahead out-of-sample forecast error, this selection strategy targets a value of λ that is optimal at a horizon (i.e. one quarter ahead) which is not the main focus of our analysis (i.e. the long-run). Second, this

¹²Following Ramey and Zubairy (2018), we exclude military spending news from the VAR in the Blanchard-Perotti robustness check, though we have verified that our findings are not affected qualitatively by this choice.

¹³Because the Blanchard-Perotti shock leads to different short-run dynamics of the government spending response, we make the shocks comparable by rescaling the IRFs in Figure 6 such that the Blanchard-Perotti shock is normalized to lead to the same cumulative increase in government spending than our baseline results over the first 16 quarters.

choice corresponds to a value of $\lambda = 0.58$, which is roughly in between the cases of relatively uninformative priors, as implied by $\lambda = 1$, and very informative priors, as implied by $\lambda = 0.1$.

As the prior mean of all autoregressive parameters with $\ell > 1$ is centered at zero and λ governs the tightness of that prior mean, varying λ effectively corresponds to ask how dogmatic one has to be to overlook long-run dynamics. To illustrate this point, it is worth noting that the exponential discounting of the lag structure embedded in the square of the parameter ℓ in equation (4) implies that the prior variance for the estimated coefficients on lag $\ell = 60$ of all variables is scaled down by a factor of $3600 = (\ell^2)/1$ in the case of the uninformative prior variance $\lambda = 1$, by $10702 = (\ell^2)/0.58^2$ for the optimal value of $\lambda = 0.58$, by $22500 = (\ell^2)/0.4^2$ for the informative scenario of $\lambda = 0.4$, by $90000 = (\ell^2)/0.2^2$ for the ‘conservative’ case of $\lambda = 0.2$ and by $360000 = (\ell^2)/0.1^2$ for the ‘dogmatic’ case of $\lambda = 0.1$. We refer to the latter values as ‘conservative’ and ‘dogmatic’ *in our context*, because they assign such a tiny prior variance around the prior mean of zero for the coefficients on longer lags that the data would need an extraordinarily large amount of information to overturn the strong belief of no effects in the long-run.¹⁴

The results of this exercise are reported in Appendix K. The columns refer to output (left) and prices (right) respectively. The first row displays the uninformative case of $\lambda = 1$, the second row refers to the informative priors implied by $\lambda = 0.4$, the third row represents the conservative specification associated with $\lambda = 0.2$ while the fourth row corresponds to the ‘dogmatic’ case of $\lambda = 0.1$. The top and bottom rows are the least surprising. The case of $\lambda = 1$ is less informative than the optimal value of 0.58 used in the rest of the paper and thus the impulse responses in the first row are very similar to those in Figure 1. At the other extreme, the ‘dogmatic’ value of $\lambda = 0.1$ in the last

¹⁴We stress *in our context* because within the quarterly VAR(4) typically used in empirical macro studies, for instance, values of 0.2 and 0.1 for λ would impose a far more moderate shrinkage (and hence would not be necessarily conservative or dogmatic) on the coefficients associated with the model most distant lag, $\ell = 4$, whose prior variance would be penalised by ‘only’ a factor of $400 = 16/0.2^2$ for $\lambda = 0.2$ and by a factor of $1600 = 16/0.1^2$ for $\lambda = 0.1$.

row implies such an extraordinarily tight variance around the prior mean of zero for most autoregressive parameters that the VAR(60) detects smaller and less significant low-frequency dynamics.

The most interesting exercises are probably the two presented in the middle rows, which show that long-run effects are still visible using the informative prior of $\lambda = 0.4$ in the second row and the conservative prior of $\lambda = 0.2$ in the third row. Unsurprisingly, decreasing exponentially the variance around the prior mean of zero for the coefficients on the more distant lags is associated with smaller estimated long-run effects when going down the rows of Figure K.1. On the other hand, it is remarkable that despite the tightness imposed by $\lambda = 0.2$, the third row still points to some non-negligible and significant effects on output beyond business cycle frequencies. This is confirmed in Figure K.2, which reports the present value multiplier across forecast horizons for each of the four values of lambda analysed in this section. Not only the profile of the multiplier across horizons is very similar to the one in Figure 2 based on the optimal value of $\lambda = 0.58$ but, even for the conservative (dogmatic) case of $\lambda = 0.2$ ($\lambda = 0.1$), the central estimate is still as large as 1.5 (1.4), with a large mass of the posterior distribution above one, around 78% (around 87%). We conclude that, unless an econometrician overwhelmingly rejects *a priori* the hypothesis of any long-run effect, the historical data suggests that government spending does have a large and significant impact on the low-frequency dynamics of output and prices.

Alternative Priors. In our baseline VAR specification, we have used “Minnesota” priors, which imply zero a-priori correlation between the coefficients of different lags and variables. In this additional robustness check, we consider two extensions that explicitly allow for a-priori correlation between the coefficients: the “sum of coefficients” prior of Doan, Litterman, and Sims (1984) and the “dummy initial observation” of Sims (1993) and Sims and Zha (1998). These priors can be used in combination with

the original Minnesota prior, and their tightness will be governed by two additional hyperparameters, μ and θ , respectively. As for the baseline case, we choose these in a hierarchical manner so as to maximize the marginal likelihood. This results in $\mu = 10$ and $\theta = 0.05$, while λ is now estimated at 0.56. The results are virtually identical to those of the baseline case and are reported in Appendix K.

Treatment of the GDP trend. In the baseline specification of Section 3, we have removed the low frequency variation in output and government spending by using the baseline estimates of GDP trend employed by Ramey and Zubairy (2018). These authors also provide another estimated measure of potential output. A popular alternative is to normalize both GDP and government spending using the one-quarter lagged values of GDP. Yet, another strategy would be to enter GDP in log-levels. Given our focus on the long-run, it may be argued that the latter could be a less restrictive assumption, as it retains the low frequency properties of the data. It would be also consistent with the treatment of the GDP deflator, which is specified in log-levels. Finally, in many monetary and fiscal VARs, both GDP and the GDP deflator are specified in log-differences.

The results from these alternative transformations of the data are reported in Appendix M and confirm, by and large, our baseline findings. On the one hand, the output response is short-lived initially and then witnesses a second significant wave after 32 quarters, despite the fact that the changes in government spending are no longer significant. On the other hand, the effects on prices are strong during the first three years after the shock and return to their historical average after a sustained period of time. In summary, different treatments of the GDP trend produce very similar results for output. The price response is somehow smaller and less persistent in two out of four GDP trend treatments; however, in all specifications, the effects of government spending on prices are highly significant for more than five years after the shock.

A different orthogonal shock. A possible concern is that the generous specification of 60 lags may introduce some spurious cycles in the impulse responses of the estimated VAR, independently of the identification scheme. Accordingly, the second wave of output response would not represent the genuine effect of government spending on GDP but the mechanical result of an over-parameterized VAR. Alternatively, it could be the case that a strong propagation mechanism in output along the lines of [Comin and Gertler \(2006\)](#) or [Beaudry et al. \(2020\)](#) are present in the data, such that *any* shock would produce the kind of highly persistent dynamics in output that we have reported in response to military spending news shocks. It follows that a similar ‘spurious’ second wave would emerge also in response to any other identified shock and thus would be misleading to infer that the long-run effects on output and prices are specific to government spending.

To assess the merit of this hypothesis, in Appendix [N](#), we present the estimated impulse responses to a ‘monetary policy shock’ using a VAR(60) and a Cholesky factorization where real GDP per-capita and the GDP deflator are ordered before the short-term interest rate. The idea behind this set of identifying restrictions is that while monetary policy responds to contemporaneous developments in output and prices, it takes at least a quarter for the effects of central bank interventions to transmit to the macroeconomy.

The reason for choosing contemporaneous zero restrictions to isolate monetary policy shocks is twofold. First, this (recursive) identification has a long tradition in monetary economics (see [Christiano et al., 2005](#), and the references contained therein). Second, 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](#)), zero restrictions have the advantage of being readily implementable in our long sample, over which neither the Fed internal forecasts nor the interest rate futures are

available. It should be noted that the purpose of this exercise is to verify whether a different orthogonal shock would produce a second wave of output effects. As such, the specific restrictions that are imposed to identify such a shock (and thus its economic interpretation) are not really crucial, as long as the identified shock is uncorrelated to other shocks.

The estimated impulse responses to a monetary policy shock are presented in Figure N.1 and they closely resemble the shape and significance typically found in the empirical monetary literature (Christiano et al., 2005). In particular, the top row shows that the structural VAR(60) estimate a short-run contraction in output and prices, peaking respectively after two and four years while returning to zero after about twenty-four quarters. More importantly, the VAR(60) detects neither a second wave of output effects nor any low-frequency movement in prices as a result of the monetary policy shock.¹⁵ We conclude that the long-run effects of government spending on output and prices that we have documented in this paper are likely to reflect a genuine feature of the data rather than an artifact of our richly parameterized model, or a systematic response of output to any type of shocks.

7 Conclusions

What are the long-run effects of government spending? Despite the resurgence in fiscal research spurred by the financial crisis of 2007-09 and the policy debate triggered by the global pandemic of 2020-22, this question has eluded empirical research. In this paper, we use 125 years of U.S. quarterly data and time series models with up to sixty lags to shed light on this issue. We argue that the combination of historical data, a

¹⁵The 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 (to isolate exogenous variation in government spending) can produce small and insignificant long-run effects.

generous lag length selection, and Bayesian shrinkage makes our framework ideal to draw inference on long-run dynamics, while retaining the ability to look also at the short-run.

We uncover four main regularities. First, the output multiplier of government spending is below one at business-cycle frequencies but is above one in the long-run. Second, while the smaller business-cycle impact can be accounted for by the crowding out of private investment, the larger effects at longer horizons appear to work through a delayed but persistent rise in productivity and innovation, leading to a subsequent expansion in private investment and a further boost in GDP. Third, there is significant heterogeneity across public spending categories: an increase in government R&D expenditure is associated with large and long-lasting effects on the economy, whereas the long-run responses of productivity, investment and GDP to an expansion in public investment and in public consumption are, respectively, smaller and insignificant. Finally, an increase in government spending of 1% GDP triggers a sustained inflation spell of around 0.15% per year for about four years, bringing the price index back to pre-shock levels in a period between five and ten years after the shock.

References

- AKCIGIT, U., J. R. GRIGSBY, T. NICHOLAS, AND S. STANTCHEVA (2022): "Taxation and Innovation in the 20th Century," *Quarterly Journal of Economics*, forthcoming.
- ASCARI, G., P. BECK-FRIIS, A. FLORIO, AND A. GOBBI (2021): "Fiscal foresight and the effects of government spending: It's all in the monetary-fiscal mix," mimeographed, University of Oxford, PIMCO and University of Milan.
- AUERBACH, A. J. AND Y. GORODNICHENKO (2012): "Measuring the Output Responses to Fiscal Policy," *American Economic Journal: Economic Policy*, 4, 1–27.
- BAEK, C. AND B. LEEB (2021): "A Guide to Autoregressive Distributed Lag Models for Impulse Response Estimations," mimeographed, University of California at Berkeley and Hong Kong University.
- BARNICHON, R. AND C. BROWNLEES (2019): "Impulse response estimation by smooth local projections," *Review of Economics and Statistics*, 101, 522–530.
- BARNICHON, R., D. DE BORTOLI, AND C. MATTHES (2022): "Understanding the Size of the Government Spending Multiplier: It's in the Sign," *Review of Economic Studies*, forthcoming.
- BARRO, R. J. (1990): "Government Spending in a Simple Model of Endogeneous Growth," *Journal of Political Economy*, 98, S103–S125.
- BARRO, R. J. AND C. J. REDLICK (2011): "Macroeconomic Effects From Government Purchases and Taxes *," *The Quarterly Journal of Economics*, 126, 51–102.
- BEAUDRY, P., D. GALIZIA, AND F. PORTIER (2020): "Putting the Cycle Back into Business Cycle Analysis," *American Economic Review*, 110, 1–47.
- BENIGNO, G. AND L. FORNARO (2017): "Stagnation Traps," *The Review of Economic Studies*, 85, 1425–1470.
- BERGEAUD, A., G. CETTE, AND R. LECAT (2016): "Productivity Trends in Advanced Countries between 1890 and 2012," *Review of Income and Wealth*, 62, 420–444.
- BLANCHARD, O. AND R. PEROTTI (2002): "An Empirical Characterization of the Dynamic Effects of Changes in Government Spending and Taxes on Output," *Quarterly Journal of Economics*, 117, 1329–68.
- CAGGIANO, G., E. CASTELNUOVO, V. COLOMBO, AND G. NODARI (2015): "Estimating Fiscal Multipliers: News From A Non-linear World," *Economic Journal*, 125, 746–776.
- CALDARA, D. AND C. KAMPS (2017): "The Analytics of SVARs: A Unified Framework to Measure Fiscal Multipliers," *Review of Economic Studies*, 84, 1015–1040.

- CENSUS (1949): *Historical statistics of the United States, 1789 - 1957*, US Department of Commerce, Bureau of the Census.
- CERRA, V., A. FATÁS, AND S. SAXENA (2022): "Hysteresis and Business Cycles," *Journal of Economic Literature*, forthcoming.
- CHODOROW-REICH, G. (2019): "Geographic Cross-Sectional Fiscal Spending Multipliers: What Have We Learned?" *American Economic Journal: Economic Policy*, 11, 1–34.
- CHONG, Y. Y. AND D. F. HENDRY (1986): "Econometric Evaluation of Linear Macroeconomic Models," *The Review of Economic Studies*, 53, 671–690.
- 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.
- CLOYNE, J., J. MARTINEZ, H. MUMTAZ, AND P. SURICO (2022): "Short-Term Tax Cuts, Long-Term Stimulus," mimeographed, University of California at Davis, London Business School and Queen Mary University of London.
- COMIN, D. AND M. GERTLER (2006): "Medium-Term Business Cycles," *American Economic Review*, 96, 523–551.
- DEL NEGRO, M. AND F. SCHORFHEIDE (2011): "Bayesian macroeconometrics," *The Oxford handbook of Bayesian econometrics*, 293, 389.
- DELEIDI, M. AND M. MAZZUCATO (2021): "Directed innovation policies and the supermultiplier: An empirical assessment of mission-oriented policies in the US economy," *Research Policy*, 50, 104151.
- DIEBOLT, C. AND K. PELLIER (2022): "Patents in the Long Run : Theory, History and Statistics," *HAL open science*, hal-02929514f.
- DOAN, T., R. LITTERMAN, AND C. SIMS (1984): "Forecasting and conditional projection using realistic prior distributions," *Econometric Reviews*, 3, 1–100.
- DONALDSON, D. AND R. HORNBECK (2016): " Railroads and American Economic Growth: A "Market Access" Approach *," *The Quarterly Journal of Economics*, 131, 799–858.
- FERNALD, J. G. (1999): "Roads to Prosperity? Assessing the Link between Public Capital and Productivity," *American Economic Review*, 89, 619–638.

- (2012): “A Quarterly, Utilization-Adjusted Series on Total Factor Productivity,” Federal Reserve Bank of San Francisco Working Paper Series 2012-19.
- FORNI, M. AND L. GAMBETTI (2016): “Government spending shocks in open economy VARs,” *Journal of International Economics*, 99, 68–84.
- FRIEDMAN, M. (1952): “Price, Income, and Monetary Changes in Three Wartime Periods,” *The American Economic Review*, 42, 612–625.
- FRIEDMAN, M. AND A. J. SCHWARTZ (1963): *A Monetary History of the United States, 1867-1960*, Princeton University Press.
- 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.
- GRANGER, C. W. J. (1969): “Investigating Causal Relations by Econometric Models and Cross-spectral Methods,” *Econometrica*, 37, 424–438.
- GROSS, D. P. AND B. N. SAMPAT (2020): “Organizing Crisis Innovation: Lessons from World War II,” Working Paper 27909, National Bureau of Economic Research.
- 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.
- HALL, G. J. AND T. J. SARGENT (2011): “Interest Rate Risk and Other Determinants of Post-WWII US Government Debt/GDP Dynamics,” *American Economic Journal: Macroeconomics*, 3, 192–214.
- HALL, R. E. (2009): “By How Much Does GDP Rise If the Government Buys More Output?” *Brookings Papers on Economic Activity*, 40, 183–249.
- HORNBECK, R. AND M. ROTEMBERG (2021): “Railroads, Market Access, and Aggregate Productivity Growth,” mimeographed, University of Chicago and New York University.
- ILZETZKI, E. (2022): “Learning by Necessity: Government Demand, Capacity Constraints, and Productivity Growth,” mimeographed, London School of Economics and Political Science.
- ILZETZKI, E., E. G. MENDOZA, AND C. A. VÉGH (2013): “How big (small?) are fiscal multipliers?” *Journal of Monetary Economics*, 60, 239–254.
- IMBS, J. M. (1999): “Technology, growth and the business cycle,” *Journal of Monetary Economics*, 44, 65–80.

- JORDÀ, Ò. (2005): "Estimation and Inference of Impulse Responses by Local Projections," *American Economic Review*, 95, 161–182.
- JORDÀ, Ò., M. SCHULARICK, AND A. M. TAYLOR (2017): "Macrofinancial history and the new business cycle facts," *NBER macroeconomics annual*, 31, 213–263.
- 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.
- LEFF YAFFE, D. (2020): "The Interstate Multiplier," mimeographed, University of California at San Diego.
- LI, D., M. PLAGBORG-MØLLER, AND C. K. WOLF (2021): "Local Projections vs. VARs: Lessons From Thousands of DGPs," mimeographed, Princeton University and University of Chicago.
- MELTZER, A. H. (2004): *A History of the Federal Reserve, Volume 1: 1913-1951*, no. 9780226520001 in University of Chicago Press Economics Books, University of Chicago Press.
- MIRANDA-AGRIPPINO, S. AND G. RICCO (2021): "The transmission of monetary policy shocks," *American Economic Journal: Macroeconomics*, 13, 74–107.
- MIRANDA-AGRIPPINO, S., G. RICCO, ET AL. (2021): *Bayesian local projections*, University of Warwick, Department of Economics.
- MONTIEL OLEA, J. L. AND M. PLAGBORG-MØLLER (2021): "Local Projection Inference Is Simpler and More Robust Than You Think," *Econometrica*, 89, 1789–1823.
- MORETTI, E., C. STEINWENDER, AND J. VAN REENEN (2019): "The Intellectual Spoils of War? Defense RD, Productivity and International Spillovers," Working Paper 26483, National Bureau of Economic Research.
- MOUNTFORD, A. AND H. UHLIG (2009): "What are the effects of fiscal policy shocks?" *Journal of Applied Econometrics*, 24, 960–992.
- MÜLLER, U. K. AND M. W. WATSON (2018): "Long-Run Covariability," *Econometrica*, 86, 775–804.
- NAKAMURA, E. AND J. STEINSSON (2014): "Fiscal Stimulus in a Monetary Union: Evidence from US Regions," *American Economic Review*, 104, 753–92.
- PAUL, P. (2017): "Historical Patterns of Inequality and Productivity around Financial Crises," Federal Reserve Bank of San Francisco Working Paper 2017-23.
- PEROTTI, R. (2004): "Public investment: another (different) look," mimeographed, Bocconi University.

- PLAGBORG-MØLLER, M. AND C. K. WOLF (2021): "Local Projections and VARs Estimate the Same Impulse Responses," *Econometrica*, 89, 955–980.
- RAMEY, V. A. (2011a): "Can Government Purchases Stimulate the Economy?" *Journal of Economic Literature*, 49, 673–85.
- (2011b): "Identifying Government Spending Shocks: It's all in the Timing," *Quarterly Journal of Economics*, 126, 1–50.
- (2019): "Ten Years after the Financial Crisis: What Have We Learned from the Renaissance in Fiscal Research?" *Journal of Economic Perspectives*, 33, 89–114.
- (2020): "The Macroeconomic Consequences of Infrastructure Investment," in *Economic Analysis and Infrastructure Investment*, National Bureau of Economic Research, Inc, NBER Chapters.
- RAMEY, V. A. AND M. SHAPIRO (1998): "Costly Capital Reallocation and the Effects of Government Spending," *Carnegie Rochester Conference on Public Policy*, 48, 145–194.
- RAMEY, V. A. AND S. ZUBAIRY (2018): "Government Spending Multipliers in Good Times and in Bad: Evidence from US Historical Data," *Journal of Political Economy*, 126, 850–901.
- ROMER, C. D. AND D. H. ROMER (2004): "A New Measure of Monetary Shocks: Derivation and Implications," *American Economic Review*, 94, 1055–1084.
- SIMS, C. A. (1993): "A nine-variable probabilistic macroeconomic forecasting model," in *Business cycles, indicators, and forecasting*, University of Chicago press, 179–212.
- SIMS, C. A. AND T. ZHA (1998): "Bayesian methods for dynamic multivariate models," *International Economic Review*, 949–968.
- UHLIG, H. (2005): "What are the effects of monetary policy on output? Results from an agnostic identification procedure," *Journal of Monetary Economics*, 52, 381–419.
- WELCH, I. AND A. GOYAL (2008): "A comprehensive look at the empirical performance of equity premium prediction," *The Review of Financial Studies*, 21, 1455–1508.

Online Appendix to
“The Long-Run Effects of Government Spending”

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

Contents

A	Data construction	2
B	Estimation algorithm	9
C	Forecast Error Variance Decomposition	10
D	Adding the long-term interest rate and private consumption	11
E	Response of Prices to Components of Public Spending	12
F	Impulse responses based on a quarterly VAR with four lags	13
G	Impulse responses based on a quarterly LP with four lags	15
H	The Multipliers at Additional Horizons	17
I	Forecast encompassing	19
J	Results based on a quarterly Bayesian LP with sixty lags	22
K	Estimates based on alternative priors	25
L	Excluding the Korean War	29
M	Impulse responses based on alternative GDP trend assumptions	30
N	Impulse Responses to Monetary Policy Shocks	32

A Data construction

The data set for our baseline estimates of the VAR and LP models comes from [Ramey and Zubairy \(2018\)](#) and contains seven variables from 1889Q1 to 2015Q5: 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. Following [Ramey and Zubairy \(2018\)](#), we scale military news, real GDP and government spending by a measure of trend GDP, estimated as a sixth-degree polynomial for the logarithm of GDP, from 1889q1 through 2015q4, excluding 1930Q1–1946Q4.

We extend the dataset in [Ramey and Zubairy \(2018\)](#) in a number of novel dimensions that we describe in turn. We first extend backwards the time series for the short-term nominal interest rate, using data from [Welch and Goyal \(2008\)](#) for the New York Fed commercial paper rate. The quarterly data on registered patents in the U.S. are sourced from IFI CLAIMS Patent Services via Google Patents Public Data.

To obtain data for private consumption and investment expenditures, we obtain annual data for consumption and investment over the sample 1870-2015 from the Macrohistory Database of [Jordà et al. \(2017\)](#). These authors provide series for real GDP, real consumption of goods (including durables), and the investment-to-output ratio, from which levels of investment can be calculated. We then interpolate the annual series to quarterly frequency in the following way: from 1919-1940, we exploit quarterly series on consumption and investment from [Gordon \(2007\)](#) to interpolate the annual series using the method in [Chow and Lin \(1971\)](#). For the period when these are not available (1889-1918 and 1941-1946), we use quarterly real GDP from [Ramey and Zubairy \(2018\)](#) to perform the interpolation. After 1947, we employ the official NIPA estimates for quarterly consumption and investment.

We also construct new time series that break down government spending into its consumption and investment components. An annual series of government investment is available from the BEA since 1914. For the period 1890-1913, we reconstruct government investment by manually transcribing data from both the *Historical Statistics of the United States* ([Census, 1949](#)) and the annual *Statistical Abstracts* published by the census. We transcribe separately data for Federal and State and Local investment. First, the *Historical Statistics*, Chapter P, p.314, provides data

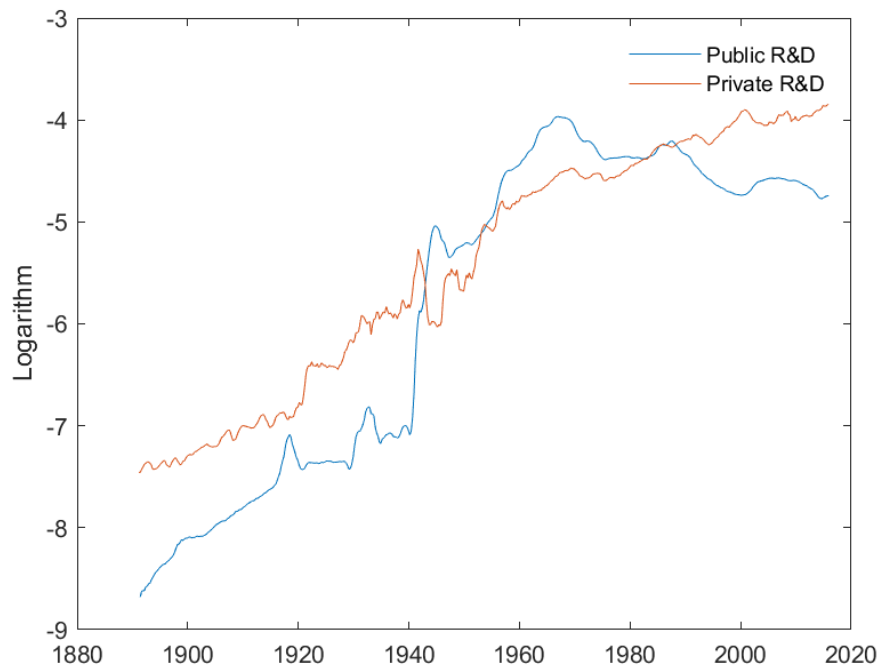
points for State and Local “capital outlays” for the years 1890, 1902, 1913. We linearly interpolate observations between these years. For Federal investments in each year between 1899 and 1921, the *Statistical Abstracts* provides detailed annual breakdowns of federal government expenditures by use over the prior ten years. We transcribe this breakdown and sum up all categories that appear to refer to investment expenditures. These include: Lighthouse Establishment, Public Buildings (Treasury Dpt.), Public Buildings (War Dpt.), Aviation (War Dpt.), Quartermasters Corps (War Dpt.), Forts and Fortifications (War Dpt.), Improving Harbors (War Dpt.), Improving Rivers (War Dpt.), Construction and Repair (Navy Dpt.), Aviation (Navy Dpt.), and Construction of Rail Roads in Alaska. We cross check that the sum of these categories is a good match to the total official figure for the years when they overlap. These estimates refer to the year ending on June 30, and therefore 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 series constructed above, we obtain an annual investment 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 quarterly series of total government spending, and finally back out government consumption as a residual.

To construct a quarterly time series for both public and private Research and Development (R&D), we rely on annual data since 1920 from the Bureau of Economic Analysis, which we interpolate to quarterly frequency using the previously described series of government and private investment, respectively. Prior to 1920, we impute quarterly values using the fitted values of a regression of public (private) R&D on four leads and for lags of both public (private) investment and patents. After 1947, we use the official quarterly data from the National Accounts. The resulting series are clearly non-stationary and enter the model in log-differences. The log-levels of the series are displayed in Figure [A.1](#).

Finally, the quarterly time series on Total Factor Productivity (TFP) has been constructed in two steps. First, we obtained annual measures of hours worked and the capital stock from [Bergeaud et al. \(2016\)](#).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

¹We are thankful to Antonin Bergeaud for sharing this data with us.

Figure A.1: PUBLIC AND PRIVATE R&D (RATIO TO POTENTIAL OUTPUT)



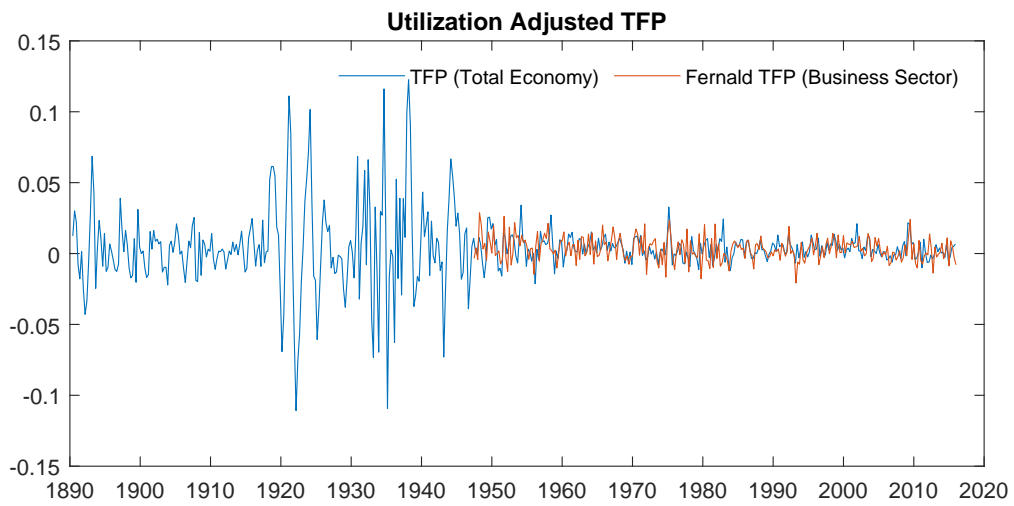
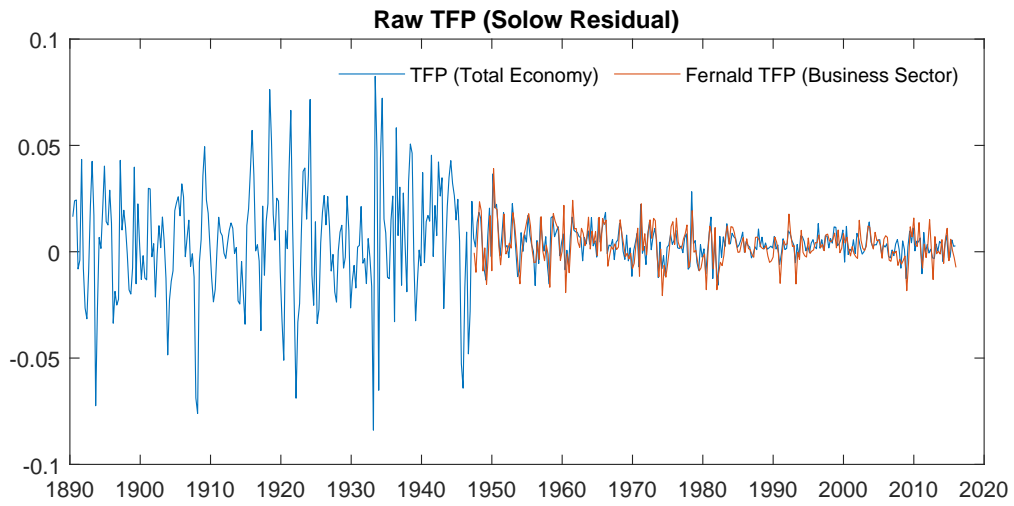
Notes. Ratio of public and private R&D to potential GDP as described in the Text.

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 \(2018\)](#). The raw TFP series is then calculated as the Solow residual using quarterly real GDP, hours worked and the capital stock, assuming a Cobb-Douglas 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 by [Imbs \(1999\)](#) (and also employed by [Paul, 2017](#); [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 A.2](#), which displays growth rates, and [Figure A.3](#), 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 [A.4](#), we report the quarterly measure of utilization adjusted TFP together with the quarterly time series of military spending news from [Ramey and Zubairy \(2018\)](#). 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.

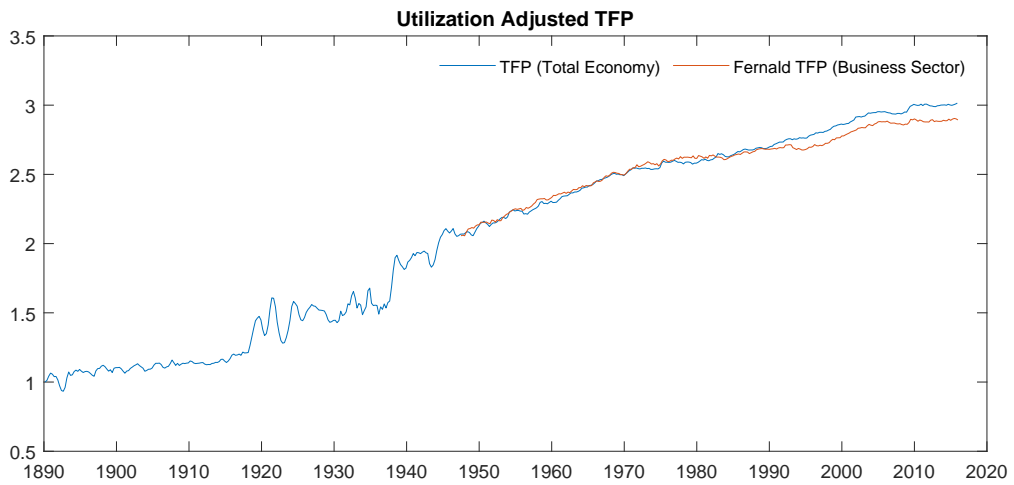
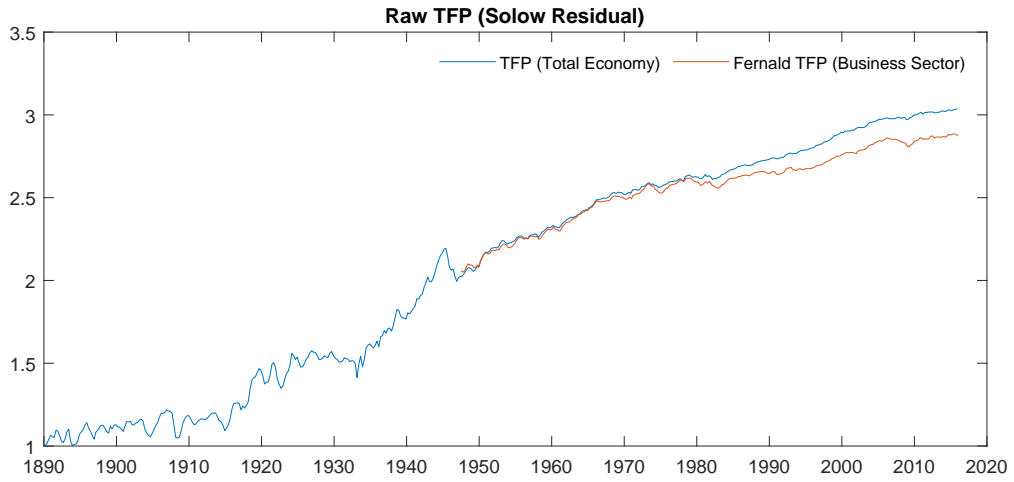
²We assume a depreciation rate of $\delta = 0.1$ per annum.

Figure A.2: RAW AND UTILIZATION ADJUSTED TFP GROWTH RATES



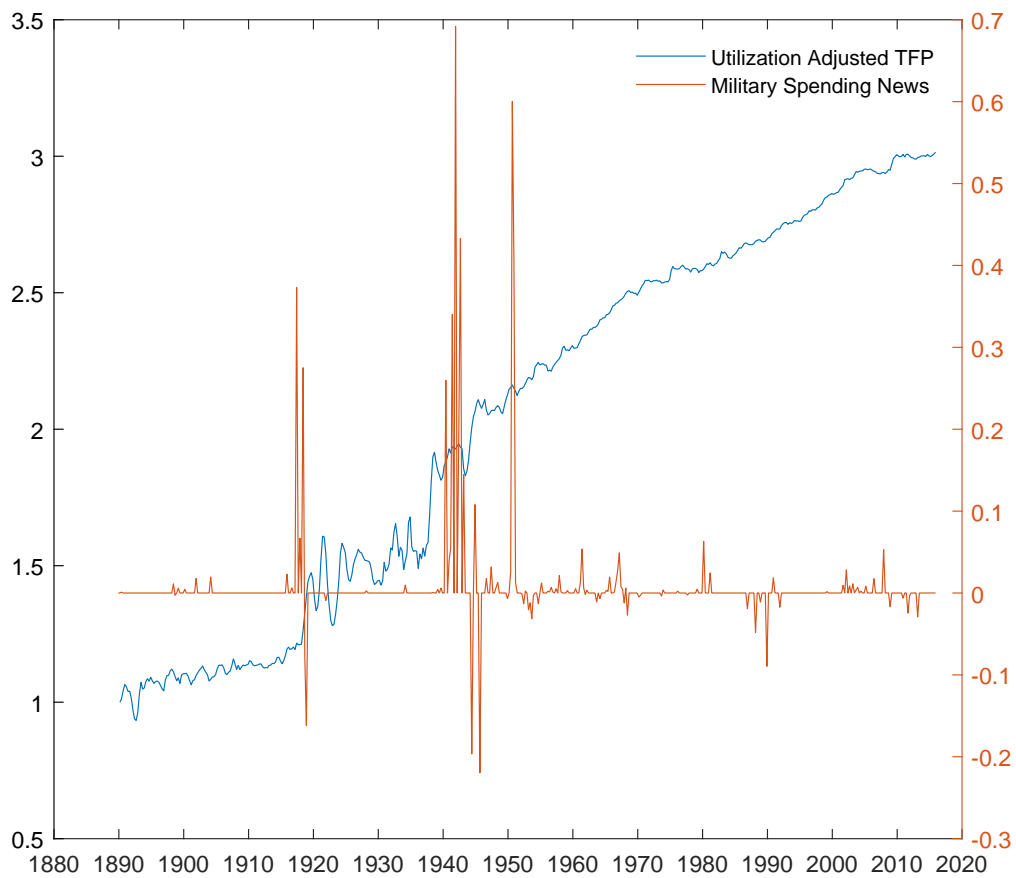
Notes. TFP Growth Rates as described in the Text. Top (bottom) row refers to the raw (utilization adjusted) TFP series.

Figure A.3: 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 A.4: 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 \(2018\)](#).

B Estimation algorithm

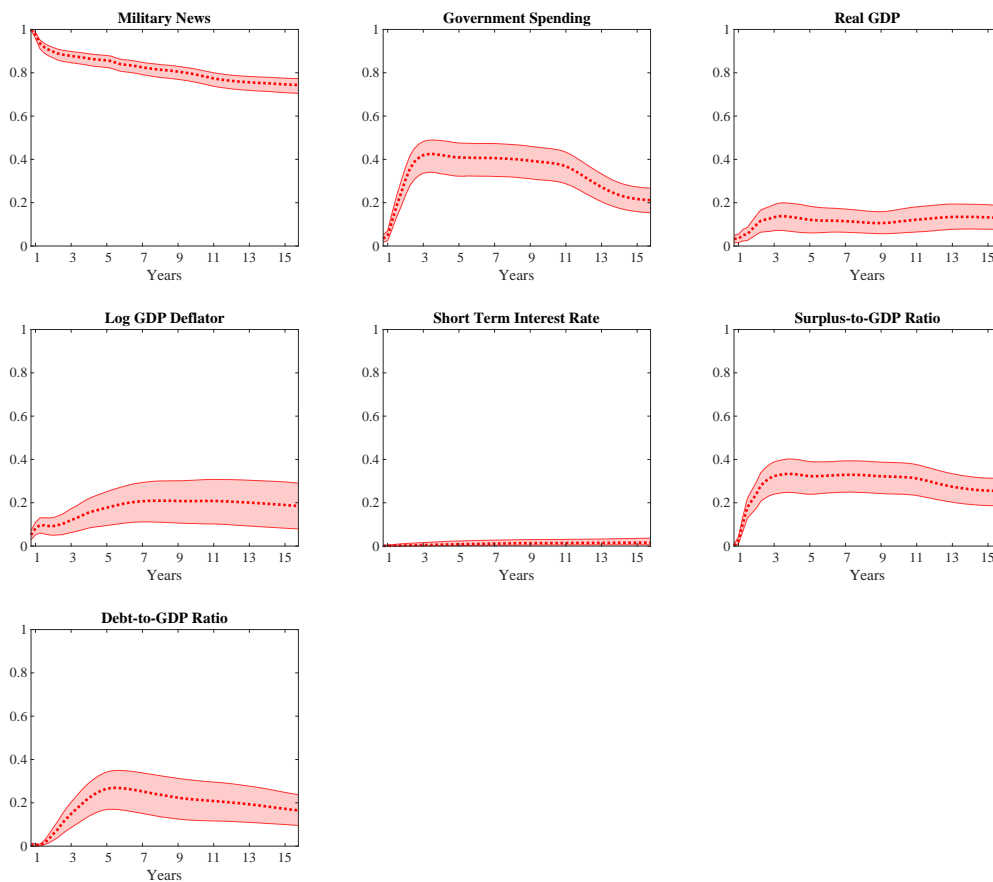
To estimate the VAR model, we can write it in matrix form as $\mathbf{Y} = \mathbf{X}\mathbf{B}' + \mathbf{U}$. Denoting T the length of the sample, n the number of variables, and p the number of lags in the VAR, $\mathbf{Y} = (\mathbf{y}'_1, \dots, \mathbf{y}'_T)'$ is a $T \times n$ matrix, $\mathbf{X}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

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

C Forecast Error Variance Decomposition

Figure C.1 reports the Forecast Error Variance Decomposition (FEVD) for the baseline results of Figure 1. The darker (lighter) shaded area represents the central 90% posterior band. The darker solid line stands for the median estimates. As can be seen from the figure, at business cycle frequencies, the military spending news shock explains about 40% of the variance of the unexpected movements in government spending, whereas it explains about 10% of the variance of real GDP and between 10% and 20% of the variance of the price level.

Figure C.1: FORECAST ERROR VARIANCE DECOMPOSITION FOR MILITARY NEWS SHOCK



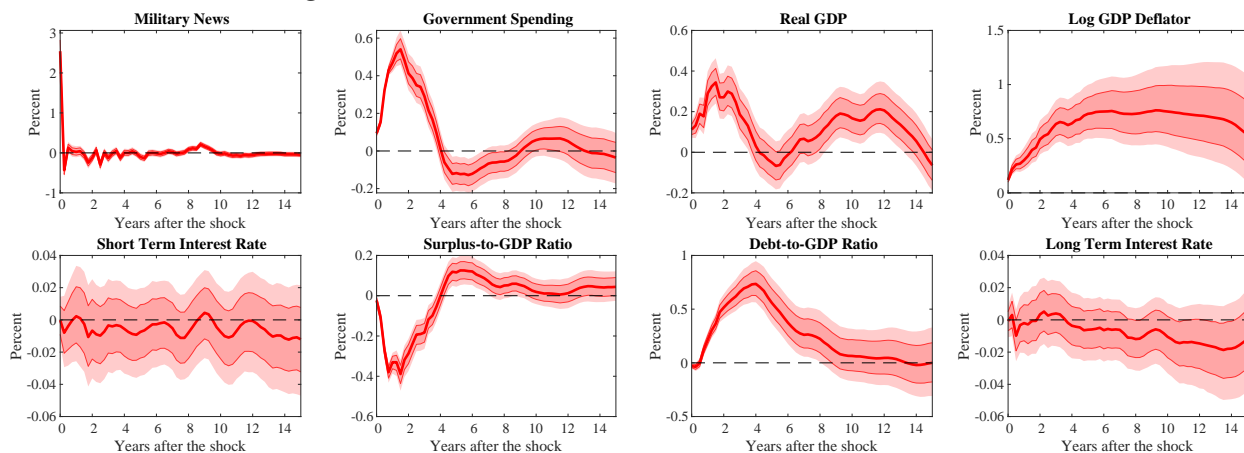
Notes. The FEVD is based on an estimated VAR with sixty lags of military spending news, government spending, real per-capita GDP, GDP deflator, short-term interest rate, fiscal deficit to GDP ratio and government debt to GDP ratio. Military spending news is ordered first in the Cholesky factorization. Output and government spending are expressed in percent of potential output as defined in Ramey and Zubairy (2018). 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, discarding explosive roots.

D Adding the long-term interest rate and private consumption

In this section, we expand the baseline seven variable VAR(60) of Section 3.1 to include, in turn, the yields on the ten-year government bond and private consumption. The results are reported in Figure D.1 and D.2, respectively.

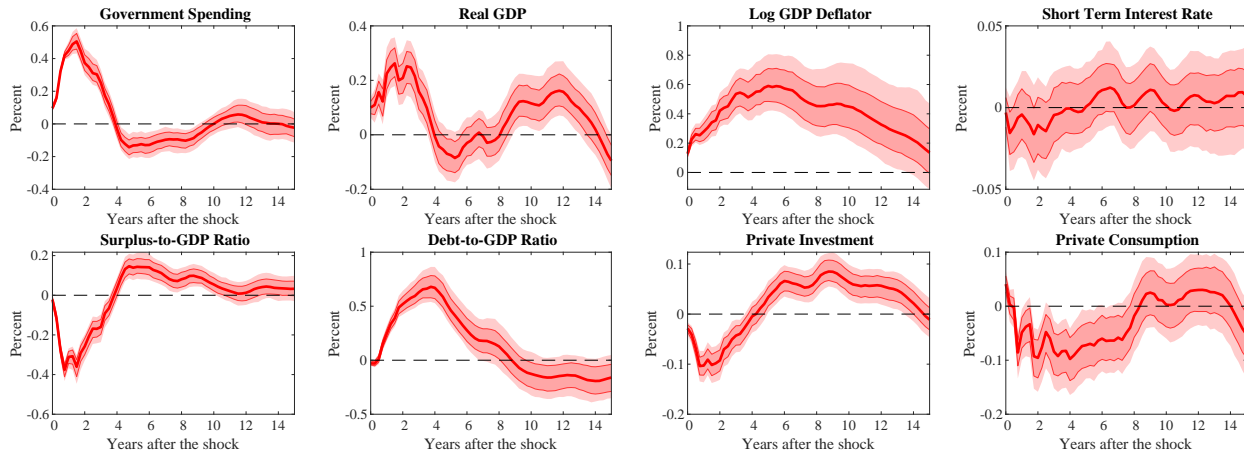
The response of the long-term interest rate in Figure D.1 is largely insignificant in response to a fiscal expansion, probably reflecting the high credibility enjoyed by the U.S. government. On the other hand, the response of consumption in Figure D.2 is negative over the first four years, consistent with the results on investment in the main text and on consumption in Ramey (2011b), but then switch to large and significant positive values. In both exercises, the estimated dynamic effects of government spending on the remaining variables of the VAR are indistinguishable from those reported in Figure 1 of the main text.

Figure D.1: ADDING THE LONG TERM INTEREST RATE



Notes. The impulse responses are based on an estimated VAR with sixty lags of military news, real GDP, GDP deflator, short-term interest rate, government spending, fiscal deficit to GDP ratio, government debt to GDP ratio, and the long term interest rate. Military spending news are ordered first in the Cholesky factorization. Output, government spending, consumption, and investment are expressed in percent of potential output as defined in Ramey and Zubairy (2018). The darker (lighter) shaded area represents the 68% (90%) HPD interval. The darker solid line stands for the median estimates. Results are based on 5000 posterior draws, discarding explosive roots.

Figure D.2: ADDING PRIVATE CONSUMPTION

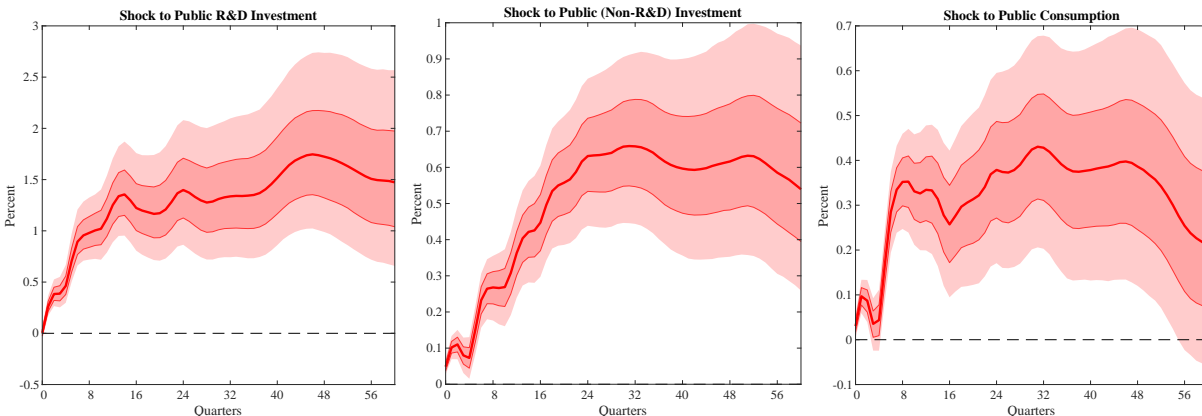


Notes. The impulse responses are based on an estimated VAR with sixty lags of military news, real GDP, GDP deflator, short-term interest rate, government spending, fiscal deficit to GDP ratio, government debt to GDP ratio, private consumption, private investment, patents and utilization adjusted TFP. Military spending news are ordered first in the Cholesky factorization. Output, government spending, consumption, and investment are expressed in percent of potential output as defined in [Ramey and Zubairy \(2018\)](#). The darker (lighter) shaded area represents the 68% (90%) HPD interval. The darker solid line stands for the median estimates. Results are based on 5000 posterior draws, discarding explosive roots.

E Response of Prices to Components of Public Spending

For completeness, in this section, we report the response of prices to each of the three government spending component shocks: public R&D, public (non-R&D) investment and public consumption.

Figure E.1: RESPONSE OF PRICES



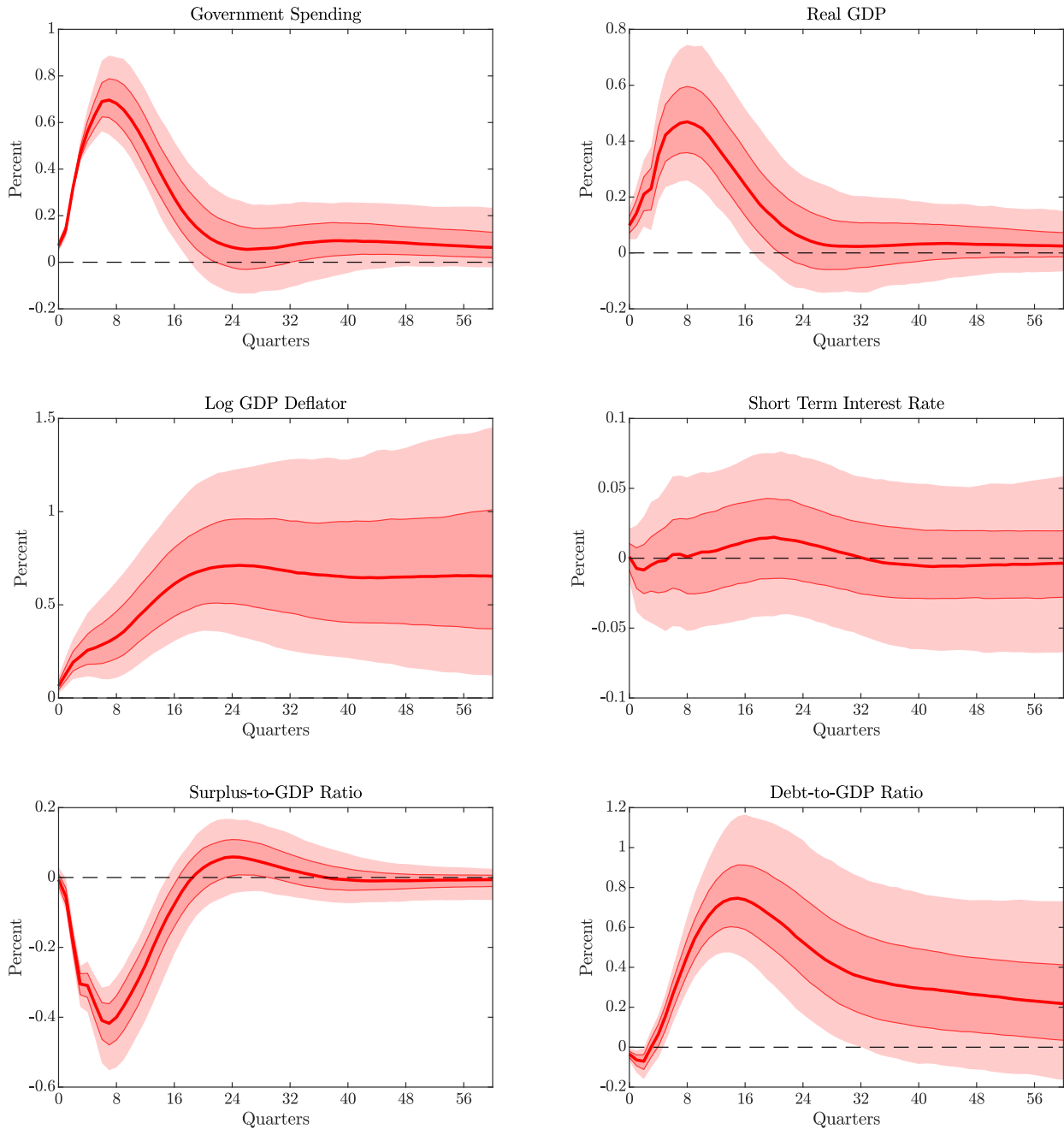
Notes. The impulse responses are based on an estimated VAR with sixty lags of real per-capita GDP, GDP deflator, short-term interest rate, fiscal deficit to GDP ratio, government debt to GDP ratio, real private investment, utilization-adjusted TFP, patents, public consumption, public non-RD investment and public RD spending. In turn, each public spending category is ordered first in the Cholesky factorization. Output, government spending, consumption, and investment are expressed in percent of potential output as defined in [Ramey and Zubairy \(2018\)](#). The darker (lighter) shaded area represents the 68% (90%) HPD interval. The darker solid line stands for the median estimates. Results are based on 5000 posterior draws, discarding explosive roots.

F Impulse responses based on a quarterly VAR with four lags

In this Appendix, we report the estimated impulse responses from a VAR that is all alike the baseline specification in Section 3, including the prior selection, but the number of lags, which in this Appendix is set to 4, rather than 60.

The main take away from Figure F.1 is that the conventional VAR(4) specification, so often used in empirical time series analysis on quarterly data, detects no significant low-frequency response for output and thus misses entirely the long-run multiplier. On the other hand, over the first few years after the shock, the VAR(4) produces estimates that are much closer to those of the VAR(60) in Figure 1. In contrast, the price dynamics implied by the VAR(4) seems much more aligned to those of the VAR(60), with the former possibly giving a more persistent response.

Figure F.1: IMPULSE RESPONSES TO MILITARY NEWS SHOCK USING VAR(4)



Notes. The impulse responses are based on an estimated VAR with four 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 and government spending are expressed in percent of potential output as defined in [Ramey and Zubairy \(2018\)](#). 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. The darker (lighter) shaded area represents the 68% (90%) HPD interval. The darker solid line stands for the median estimates. Results are based on 5000 posterior draws, discarding explosive roots.

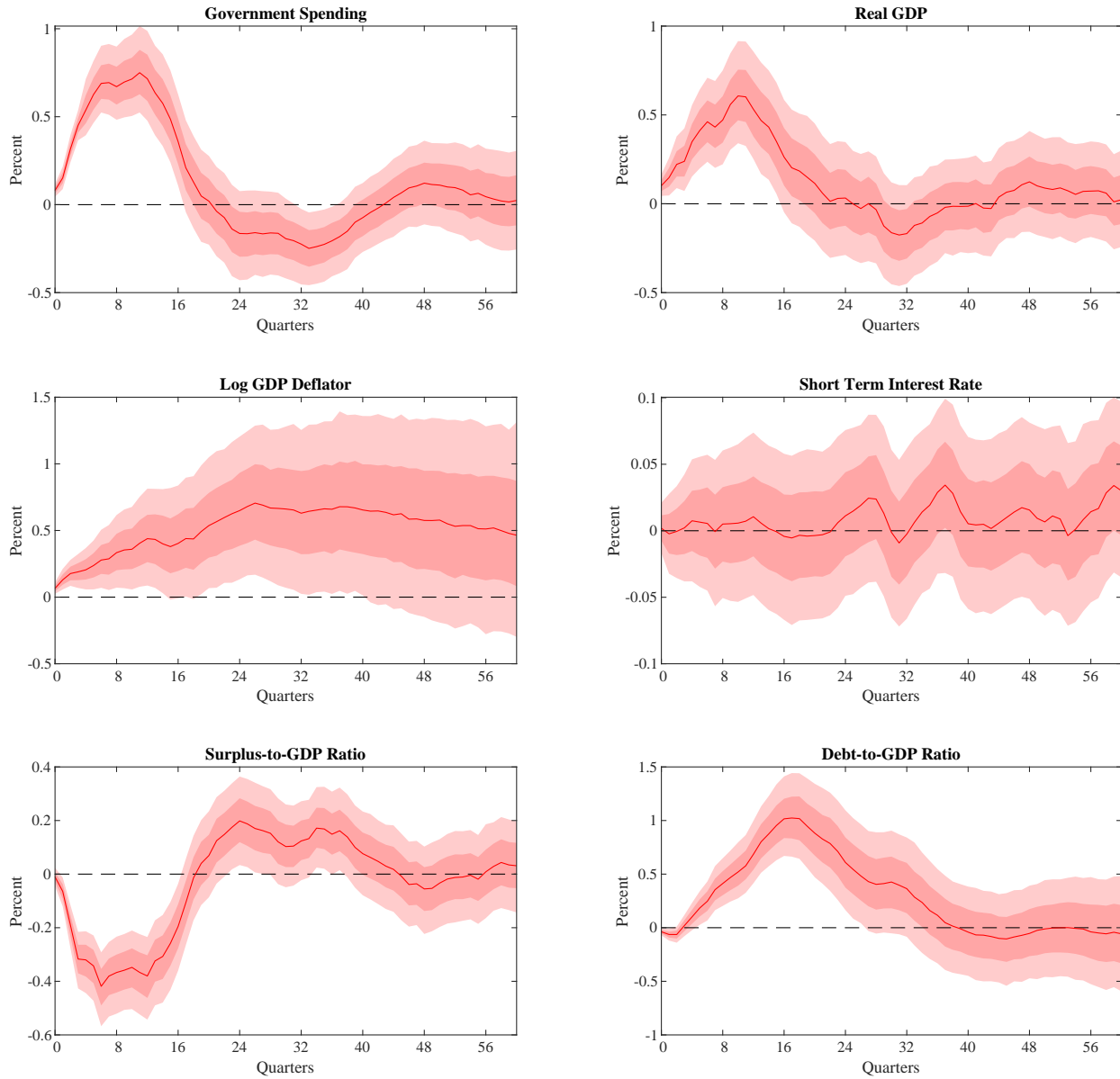
G Impulse responses based on a quarterly LP with four lags

In this section, we use the same variables and set of priors used in the previous appendix to estimate the effects of government spending on output and prices using local projections. Consistently with the VAR in the previous appendix and a large part of the empirical macro literature using quarterly data, we use *four* lags of all seven variables in our data as controls, in a specification which we refer to as LP(4).

The purpose of this exercise is twofold. On the one hand, it provides an alternative specification to the VAR(4) of the previous section, thereby feeding into the influential literature on VARs vs LPs initiated by [Plagborg-Møller and Wolf \(2021\)](#) and [Li et al. \(2021\)](#). Second, it allows us to compare the estimates in the last rows of [Table 1](#) and [Figure 6](#) based on a LP(60), with the estimates one would obtain setting instead to 4 the lag length of the controls (i.e. LP(4)), as often done in the empirical macro literature using quarterly data.

The findings from this exercise are collected in [Figure G.1](#), which essentially convey two main messages. First, also the LP(4), exactly like the VAR(4), fails to detect the large and significant long-run effects of government spending on output that dominate the estimated based on the LP(60) and the VAR(60) in the main text. Second, and again in line with the results of the VAR(4), the impulse response of prices (and indeed any other variables) from the LP(4) are very much in line with those obtained using a specification with many more lags.

Figure G.1: IMPULSE RESPONSES TO MILITARY NEWS SHOCK USING LP(4)



Notes. The impulse responses are based on an estimated LP with four 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. Output and government spending are expressed in percent of potential output as defined in [Ramey and Zubairy \(2018\)](#). 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. The darker (lighter) shaded area represents the 68% (90%) HPD interval. The darker solid line stands for the median estimates. Results are based on 5000 posterior draws, discarding explosive roots.

H The Multipliers at Additional Horizons

The empirical literature on the dynamic effects of government spending has presented a plethora of output multipliers that have been typically estimated at different short-run and business cycle horizons across papers. While the main focus of our analysis is to provide novel evidence on the overlooked long-run multiplier, in Table H.1 of this Appendix we complement the analysis in the main text by computing the output multiplier of government spending at the conventional, shorter horizons of one quarter (Panel A), two years (Panel B) and four years (Panel C) used in earlier studies. A main take away from this exercise is that, by providing a systematic analysis of the output multiplier across forecast horizons (within each and every specification), we are able to span the whole range of estimates available in the empirical literature. This suggests that the seemingly conflicting results in earlier work may simply reflect the fact that different studies focus on different forecast horizons.

The left portion of each panel of Table H.1 refers to VAR specifications whereas the sections on the right correspond to LP models. Each row represents a different number of lags for the relevant model in that row, ranking from a minimum of 4 lags to a maximum of 60 lags. The columns (from left to right) stand for the 5th, 50th and 95th of the posterior distribution of the multiplier of interest for the specification in each row whereas the last column in each section, headed with $\mathcal{M} > 1$, records the share of draws for which the multiplier at the horizon in that panel and for that specification is above one.

Two main results emerge from Table H.1. First, independently of whether we use VARs or LPs and independently of the lag length selection, the entries within each panel (i.e. within each of the shorter forecast horizons in the table) are fairly similar to each other. This suggests that omitting higher lags in either VAR or LP specifications is inconsequential for the estimates of the short-run multiplier, despite the evidence shown in the main text about the sizable bias that omitting those lags produce when estimating the long-run multiplier. Second, and again very robustly across models and specifications, the impact multiplier one quarter head is about twice as large as the output multipliers at two and four years horizons, such that the share of draws for which the posterior distribution of the multiplier is above one range from 64% to 88% after one quarter but is no larger than 3% and 8% at the frequencies of two and four years, respectively.

Table H.1: THE PRESENT VALUE MULTIPLIER, \mathcal{M} , ACROSS FORECAST HORIZONS

Panel A. Multiplier at 1-quarter horizon

No. lags (p)	Vector AutoRegressions (VAR)				Bayesian Local Projections (LP)			
	5^{th}_{pct}	50^{th}_{pct}	95^{th}_{pct}	$\mathcal{M} > 1$	5^{th}_{pct}	50^{th}_{pct}	95^{th}_{pct}	$\mathcal{M} > 1$
4	0.71	1.34	2.43	80%	0.61	1.26	2.32	73%
10	0.78	1.39	2.37	84%	0.68	1.39	2.68	80%
20	0.80	1.35	2.15	85%	0.64	1.29	2.35	75%
30	0.70	1.21	1.92	74%	0.49	1.17	2.27	64%
40	0.79	1.30	2.05	82%	0.54	1.30	2.77	72%
50	0.85	1.31	2.00	85%	0.60	1.35	3.03	76%
60	0.89	1.35	2.06	88%	0.58	1.40	3.35	76%

Panel B. Multiplier at 2-year horizon

No. lags (p)	Vector AutoRegressions (VAR)				Bayesian Local Projections (LP)			
	5^{th}_{pct}	50^{th}_{pct}	95^{th}_{pct}	$\mathcal{M} > 1$	5^{th}_{pct}	50^{th}_{pct}	95^{th}_{pct}	$\mathcal{M} > 1$
4	0.41	0.66	0.91	2%	0.54	0.67	0.80	0%
10	0.48	0.71	0.94	3%	0.58	0.70	0.84	0%
20	0.51	0.71	0.92	1%	0.55	0.67	0.80	0%
30	0.41	0.60	0.79	0%	0.47	0.60	0.74	0%
40	0.42	0.64	0.85	0%	0.54	0.67	0.81	0%
50	0.49	0.67	0.86	0%	0.57	0.70	0.85	0%
60	0.48	0.65	0.83	0%	0.55	0.67	0.80	0%

Panel C. Multiplier at 4-year horizon

No. lags (p)	Vector AutoRegressions (VAR)				Local Projections (LP)			
	5^{th}_{pct}	50^{th}_{pct}	95^{th}_{pct}	$\mathcal{M} > 1$	5^{th}_{pct}	50^{th}_{pct}	95^{th}_{pct}	$\mathcal{M} > 1$
4	0.46	0.71	0.98	4%	0.63	0.72	0.83	0%
10	0.57	0.78	1.00	5%	0.63	0.73	0.83	0%
20	0.56	0.76	0.95	2%	0.60	0.70	0.80	0%
30	0.49	0.67	0.86	0%	0.59	0.71	0.84	0%
40	0.51	0.76	1.03	8%	0.68	0.81	0.96	2%
50	0.56	0.78	1.01	6%	0.62	0.75	0.90	0%
60	0.50	0.70	0.90	1%	0.56	0.66	0.77	0%

Notes: The VAR and LP specifications use seven variables: 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 of the VAR. The definition of variables and the definition of the present value multiplier over 60 quarters horizon, \mathcal{M} , follow [Ramey and Zubairy \(2018\)](#) and [Mountford and Uhlig \(2009\)](#). Results are based on 5000 posterior draws, discarding explosive roots. Each row refers to the estimates of a different specification of either the VAR(p) or the LP using p lags of all variables. The number of lags, p , selected in each specification is reported in the first column. The columns 5^{th}_{pct} , 50^{th}_{pct} and 95^{th}_{pct} present the 5^{th} , 50^{th} and 95^{th} percentiles of the posterior distribution of \mathcal{M} . The columns $\mathcal{M} > 1$ report the share of posterior draws for which the cumulated response of GDP over the cumulated response of government spending is larger than one.

I Forecast encompassing

In this section, we look at one possible metrics along which to compare the predictive accuracy of our VAR(60) with the predictive accuracy of the popular VAR(4) used in the empirical macro literature. This is the encompass strategy in [Chong and Hendry \(1986\)](#), who recommend to estimate two competing models on artificial data generated using the estimates of the other specification in order to evaluate which model is able to produce forecasts that encompass the forecasts of the other model.

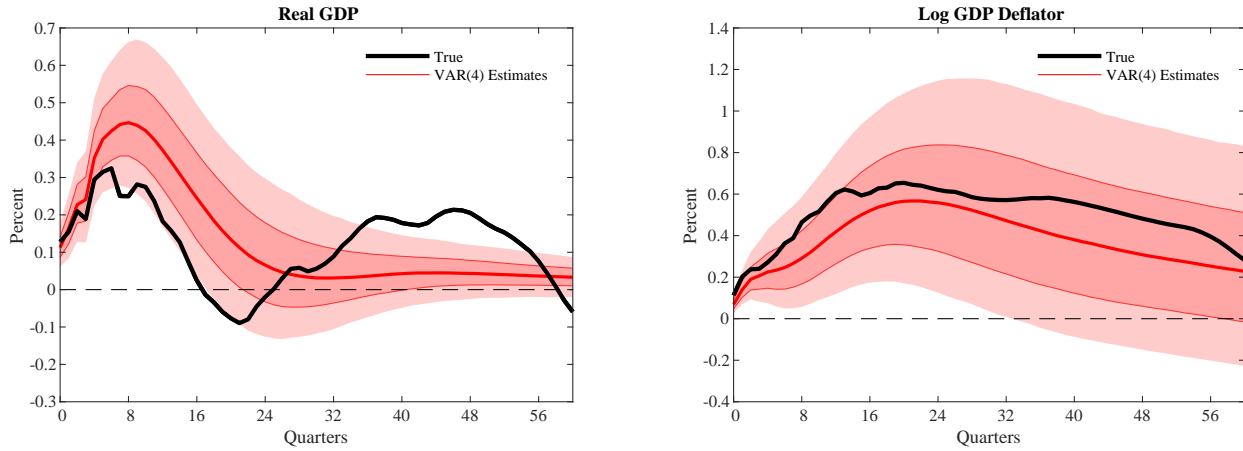
We proceed in two symmetric steps. In the first step, we generate one set of artificial data using the point estimates of the VAR (60) on actual data and then fit on these artificial data a VAR(4) specification. In the second step, we do the reverse and generate another set of artificial data, using this time the point estimates of the VAR(4) on actual data, and then fit on these artificial data a VAR(60). If the impulse responses of the VAR(60) are within the credible sets of estimates for the impulse responses of the VAR(4) when the data generating process is the VAR(4) *but* the impulse responses of the VAR(4) are outside the credible sets of estimates for the impulse responses of the VAR(60) when the data generating process is the VAR(60), then we conclude that (the forecasts of) the VAR(60) encompass (the forecasts of) the VAR(4) but the VAR(4) does not encompass the VAR(60).

The findings from this exercise are reported in [Figure I.1](#). The top row presents the VAR(4) estimates when the data generating process is the VAR(60) whereas the bottom row displays the VAR(60) estimates when the data generating process is the VAR(4). The impulse responses of GDP are in the first column while the impulse responses of prices are in the second column. The top left panel reveals that the VAR(4) estimates (in red) have hard time to match the true impulse response of output (in black) when the data generating process is the VAR(60). This is true not only at most frequencies beyond 32 quarters (i.e. in the long-run) when the true long-run effects on output are well above the small and insignificant effects estimated by the VAR(4), but also, though at a lesser extent, in years 2 and 3 when the VAR(4) estimates are now significantly larger than in the data generating process. In sharp contrast, the bottom left panel makes it clear that, when the data generating process is the VAR(4), the true impulse responses are always inside the credible sets of estimates based on the VAR(60), especially at lower frequencies.

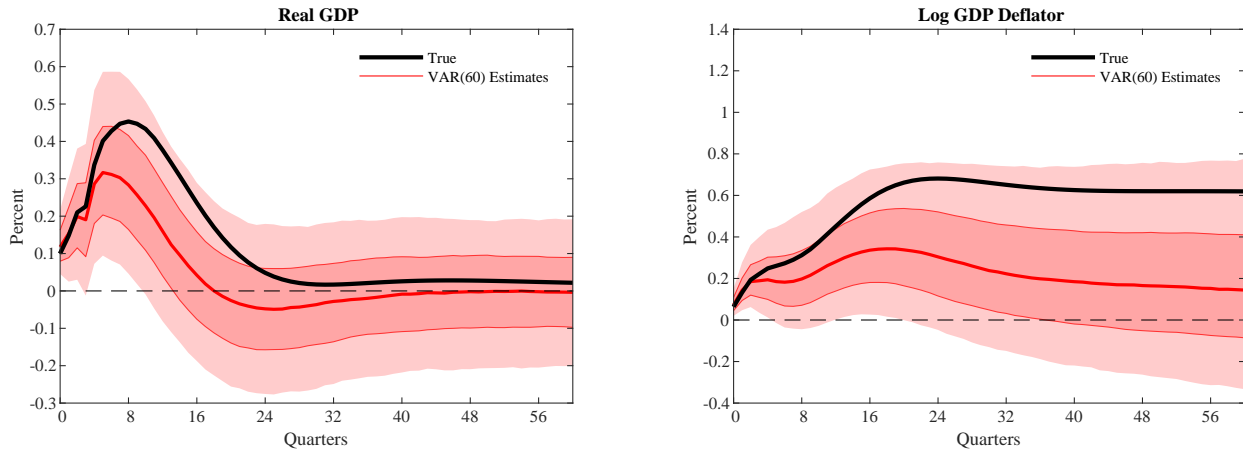
On the other hand, the two models fair equally well on prices as both credible sets include the true impulse responses in the data generating process. We conclude that the VAR(60) is able to detect small and insignificant long-run effects on output when, indeed, there are none in the data generating process whereas the VAR(4) is unable to pick up any long-run effect on output when, in fact, these are large and significant in the data generating process. In the language of [Chong and Hendry \(1986\)](#), one can say that the forecasts of the VAR(60) encompass those of the VAR(4) but the converse is not true. We interpret these findings as further suggestive evidence that a higher number of lags is desirable to draw reliable inference on both the long-run and the short-run.

Figure I.1: ESTIMATES FROM ALTERNATIVE DGPs AND VAR SPECIFICATIONS

(a) VAR(4) estimates when DGP is VAR(60)



(b) VAR(60) estimates when DGP is VAR(4)



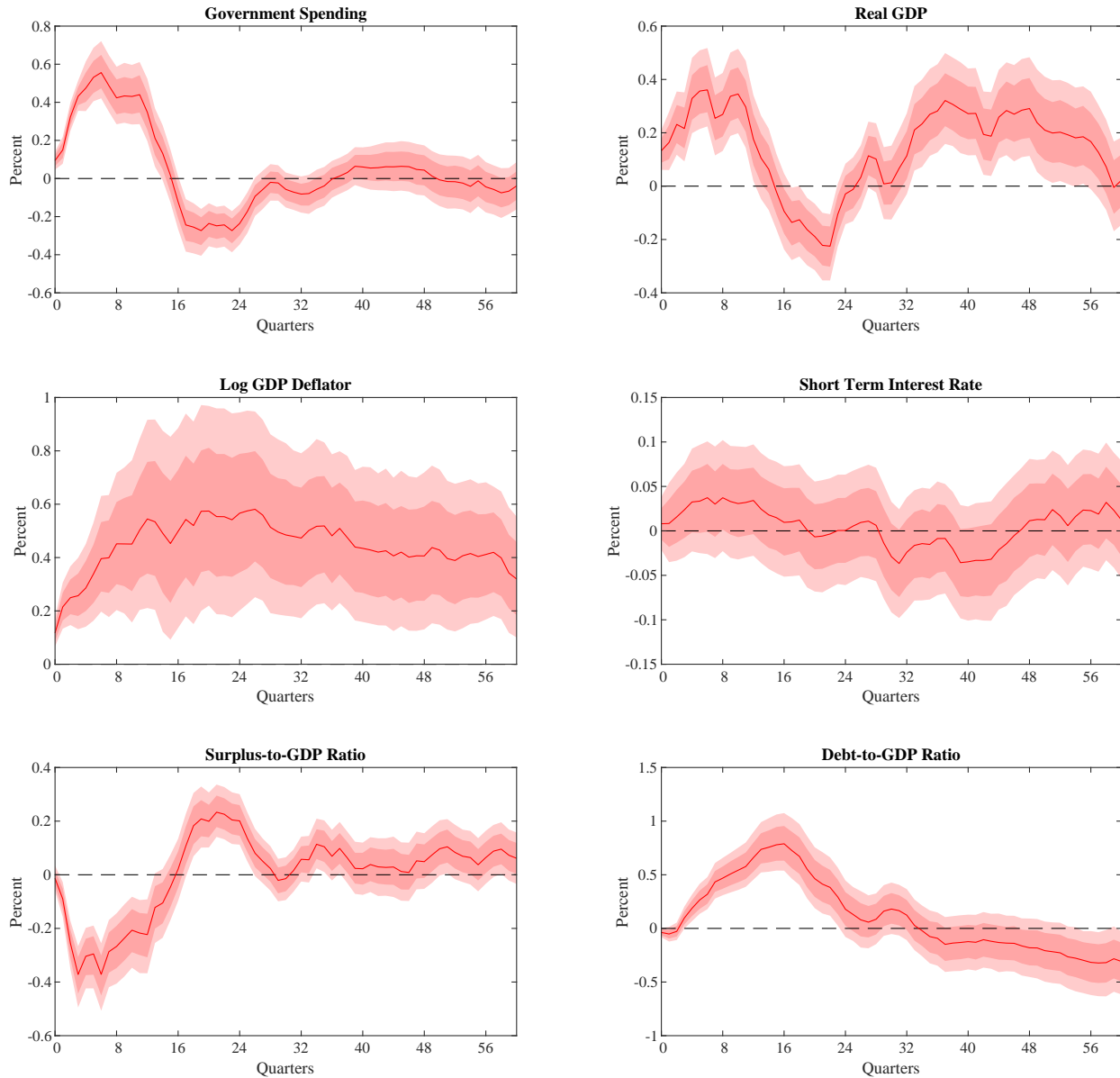
Note: The red solid lines represent the median posterior responses. The darker shadow area represents the 68th posterior credible intervals, while the lighter shadow are represents the 95th posterior credible intervals. The black solid line refers to the true impulse response in the data generating process, which is a VAR(60) in the top row and a VAR(4) in the bottom row estimated on actual data on military spending news, government spending, real GDP per-capita, GDP deflator, short-term interest rate, fiscal deficit to GDP ratio and government debt to GDP ratio.

J Results based on a quarterly Bayesian LP with sixty lags

In this section, we report the full set of results associated with the estimates of the LP(60), whose impulse responses for real GDP and prices have been reported among the sensitivity checks of Figure 6 (top row). In particular, in Figure J.1 we present also the impulse response for government spending, the short-term interest rate, the fiscal surplus and the public debt-to-gdp ratio. All estimates are fairly similar to those based on the VAR(60), consistent with the finding in [Plagborg-Møller and Wolf \(2021\)](#) that VARs and LPs estimate the same impulse response (in large samples) whenever the span of the forecast horizon is as large as the number of lags used to estimate each model.

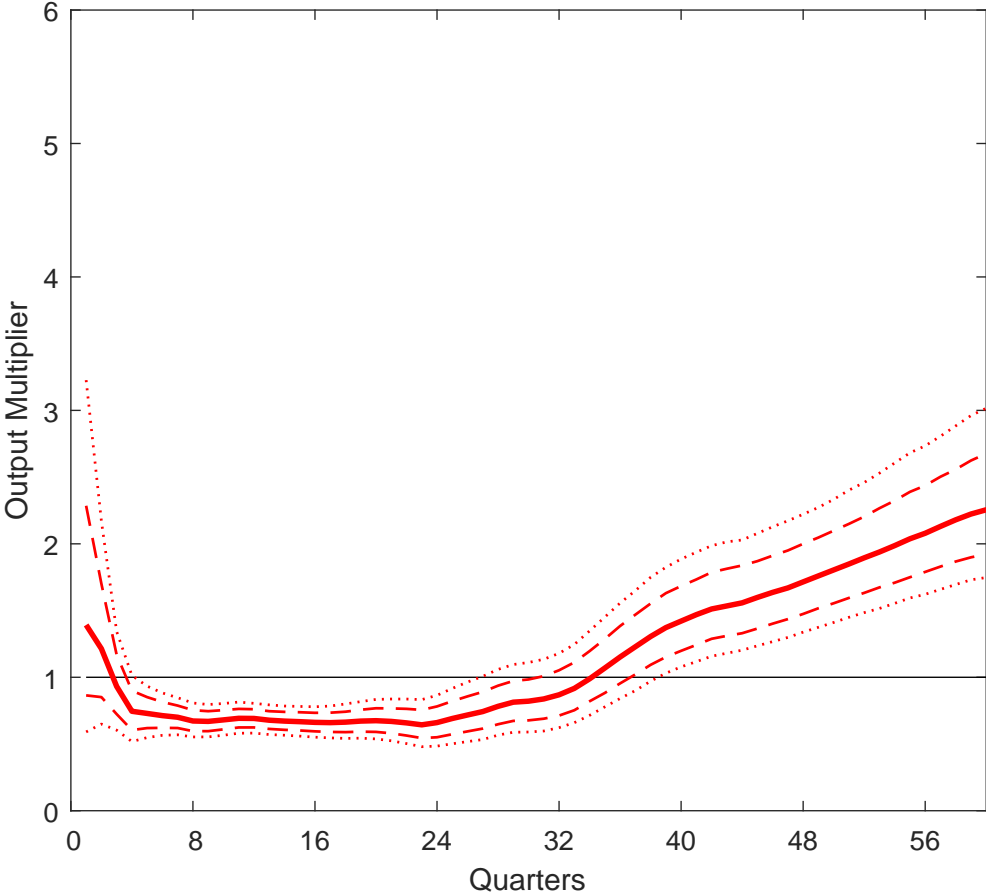
Figure J.2 is the counterpart of Figure 2 and show the present value multiplier based on the estimated LP(60) as a function of the forecast horizon. Very much in line with the findings from the VAR(60) in Figure 2, the estimates in Figure J.2 confirms the pervasive heterogeneity across forecast horizons. On impact the multiplier is large, with most of the posterior distribution mass above one. At business-cycle frequencies, between 6 quarters and 8 years, the multiplier is smaller and most of the time significantly below one. But in the long-run (i.e. beyond the business cycle frequencies of 32 quarters), the present value multiplier becomes significantly larger than one at most horizons, with a central estimate of 2.26 sixty quarters ahead.

Figure J.1: IMPULSE RESPONSES TO MILITARY NEWS SHOCK USING LP(60)



Notes. The impulse responses are based on an estimated Bayesian LP 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. Contemporaneous military news is taken to be the shock of interest. Output and government spending are expressed in percent of potential output as defined in [Ramey and Zubairy \(2018\)](#). 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. The darker (lighter) shaded area represents the 68% (95%) HPD interval. The darker solid line stands for the median estimates. Results are based on 5000 posterior draws.

Figure J.2: THE PRESENT VALUE MULTIPLIER ACROSS FORECAST 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 using the interest rate adjustment for discounting as in [Mountford and Uhlig \(2009\)](#). The estimates are based on a Bayesian LP 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. Contemporaneous military news is taken to be the shock of interest. Output and government spending are expressed in percent of potential output as defined in [Ramey and Zubairy \(2018\)](#). The broken (dotted) lines represent the central 68% (90%) posterior band. The solid line stands for the median estimates. Results are based on 5000 posterior draws.

K Estimates based on alternative priors

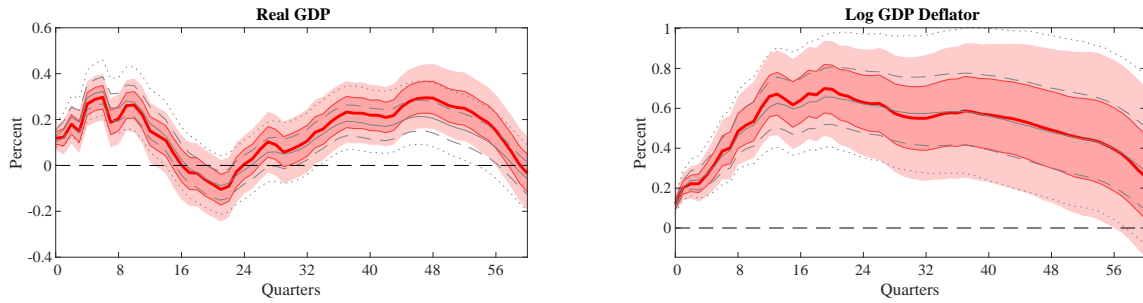
In this section, we present the impulse responses of output and prices based on four variants of the VAR(60) estimated in the main text. The four versions differ in the tightness of the prior variance around the mean of zero for all autoregressive coefficients after the first lag. The results are reported in Figure K.1, whose rows refer to progressively tighter specifications, from the uninformative priors associated with $\lambda = 1$ (see equation (4)) in the first row, to the informative scenario of $\lambda = 0.4$ in the second row, the conservative value of $\lambda = 0.2$ in the third row and the dogmatic case of $\lambda = 0.1$ in the fourth row.

Three main results emerge from this sensitivity analysis. First, progressively tighter priors on the autoregressive coefficients of the VAR(60) going down the rows of Figure K.1 are associated with progressively smaller long-run effects. Second, despite this progressively increasing tightness, it is still the case that government spending has non-negligible and significant long-run effects on output and prices even in the conservative specification of $\lambda = 0.2$ in the third row. Third, one has to hold an overwhelmingly tight belief that longer lags are actually zero (as in the last row of $\lambda = 0.1$) to find smaller and less significant long-run effects of government spending on output. But even in this dogmatic case, Figure K.2 reveals that most of the mass of the posterior distribution of the present value multiplier is still above one in the long-run.

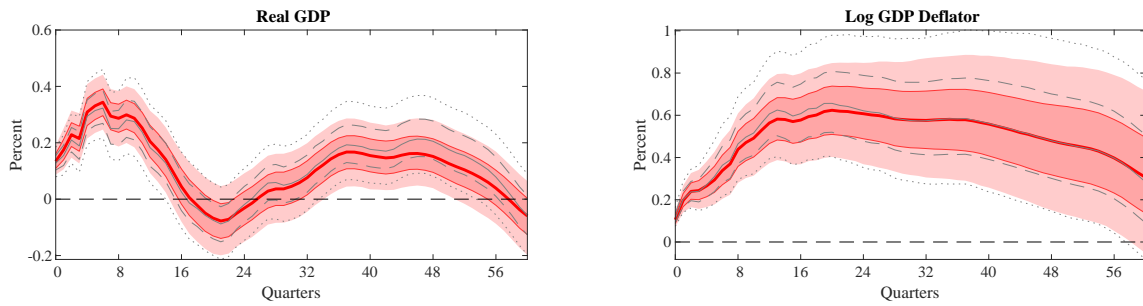
Finally, Figure K.3 reports as an additional robustness check two extensions that explicitly allow for a priori correlation between the coefficients, the “sum of coefficients” prior of Doan, Litterman, and Sims (1984) and the “dummy initial observation” of Sims (1993). These priors can be used in combination with the original Minnesota prior, and their tightness will be governed by two additional hyperparameters, μ and θ , respectively. As was the case before, we choose these in a hierarchical manner in order to maximize the marginal likelihood, resulting in $\mu = 10$ and $\theta = 0.05$, with λ now estimated at 0.56. It is noteworthy that maximization of the marginal likelihood leads to a value of μ which implies that the sum of coefficients prior is essentially uninformative, whereas the dummy initial observation is quite tight and the Minnesota prior retains approximately the same tightness. As can be seen from the figure the results are essentially identical to those of the base case and reported in Appendix K.

Figure K.1: IMPULSE RESPONSE FUNCTIONS UNDER ALTERNATIVE TIGHTNESS OF PRIOR

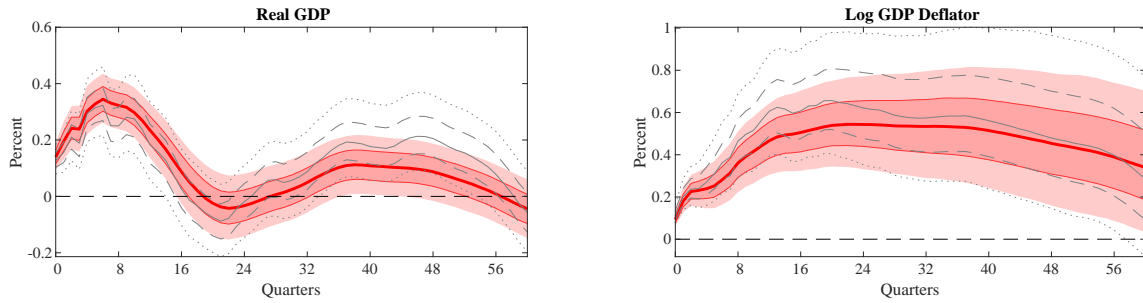
(a) $\lambda = 1$



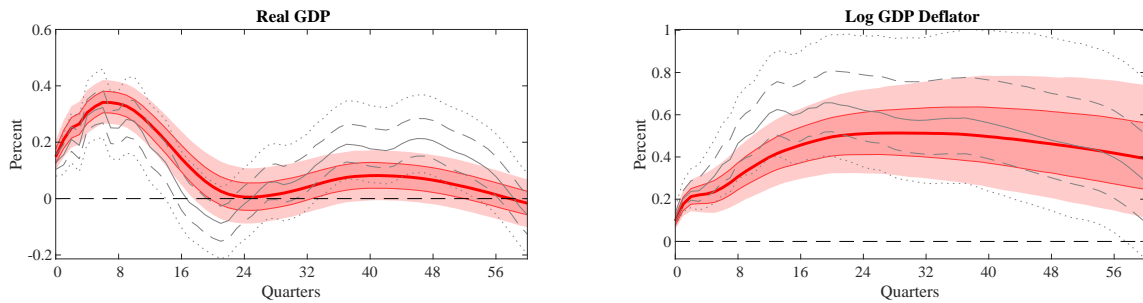
(b) $\lambda = 0.4$



(c) $\lambda = 0.2$

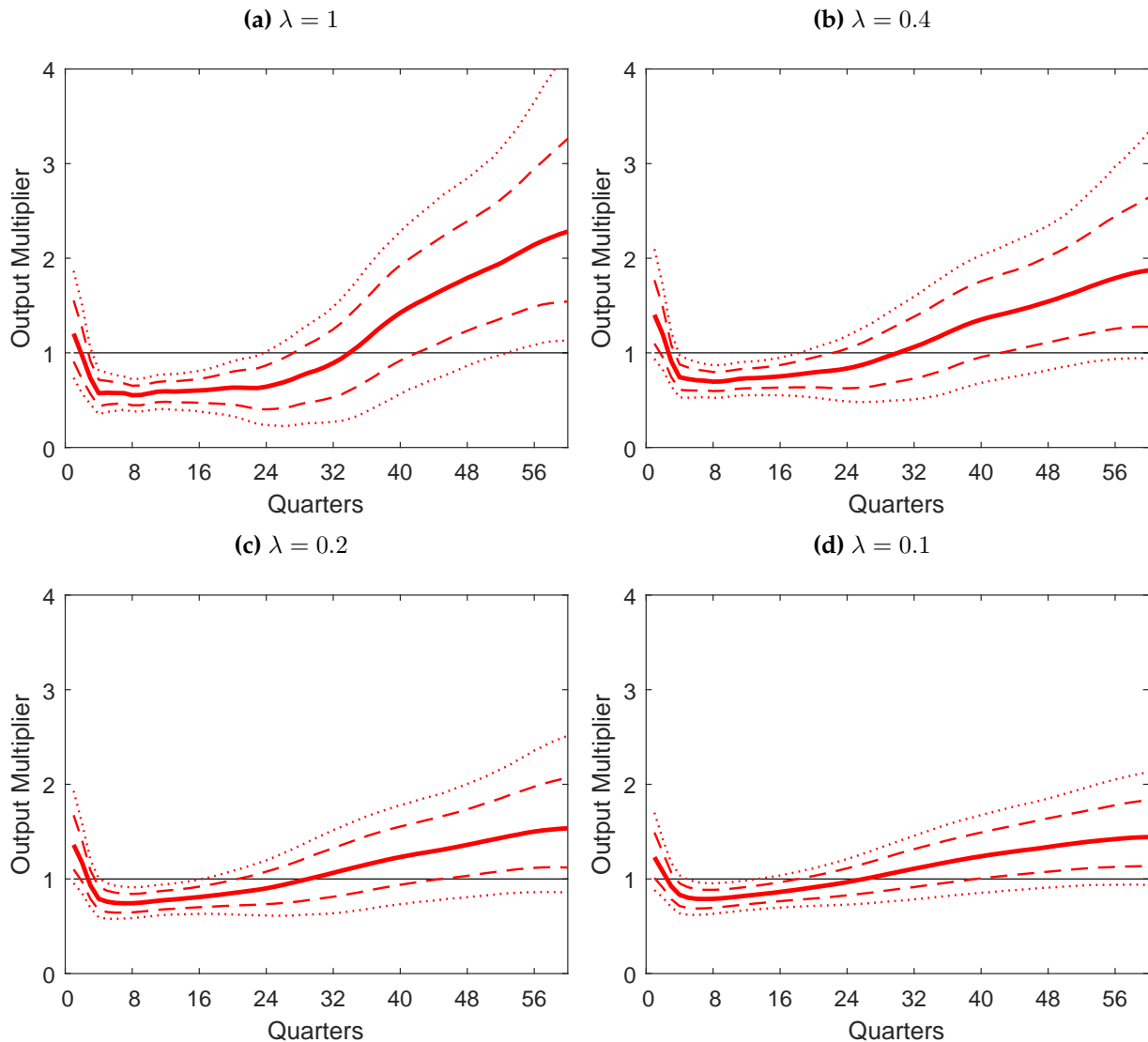


(d) $\lambda = 0.1$



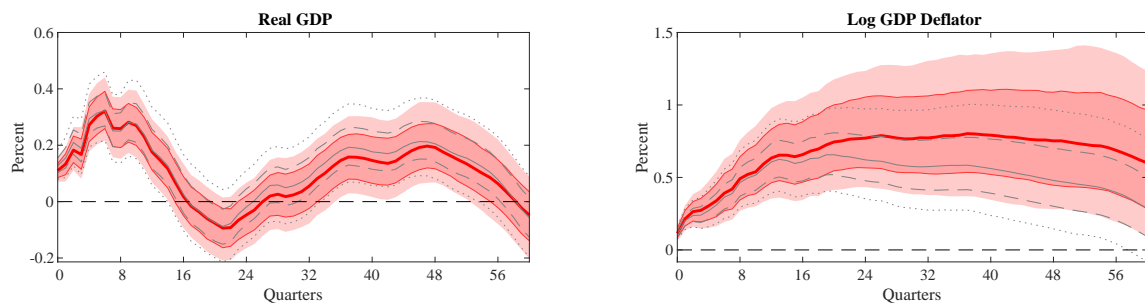
Note: The solid lines represent the median posterior response. The darker (lighter) shadow area represents the 68th (95th) HPD interval. All specifications use sixty lags of military spending news, government spending, GDP, GDP deflator, treasury bills, deficit to GDP ratio and debt to GDP ratio as in the baseline model. In each row, the parameter λ that governs the tightness of the priors 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. Solid, broken and dotted lines in light grey represent the credible sets and median estimates of the impulse responses of output and prices based on the optimal prior of $\lambda = 0.58$ that we use as baseline specification in the main text.

Figure K.2: PRESENT VALUE OUTPUT MULTIPLIER UNDER ALTERNATIVE PRIOR TIGHTNESS



Note: 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 using the interest rate adjustment for discounting as in [Mountford and Uhlig \(2009\)](#). The estimates are based on a 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, as in the baseline model. Military spending news is ordered first in the Cholesky factorization. Output and government spending are expressed in percent of potential output as defined in [Ramey and Zubairy \(2018\)](#). In each row, the parameter λ that governs the tightness of the priors 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 to 0.1 in the bottom row. The red broken (dotted) lines represent the 68% (90%) HPD interval. The red solid line stands for the median estimates. Results are based on 5000 posterior draws, discarding explosive roots.

Figure K.3: IMPULSE RESPONSE FUNCTIONS UNDER ALTERNATIVE SPECIFICATION OF PRIOR

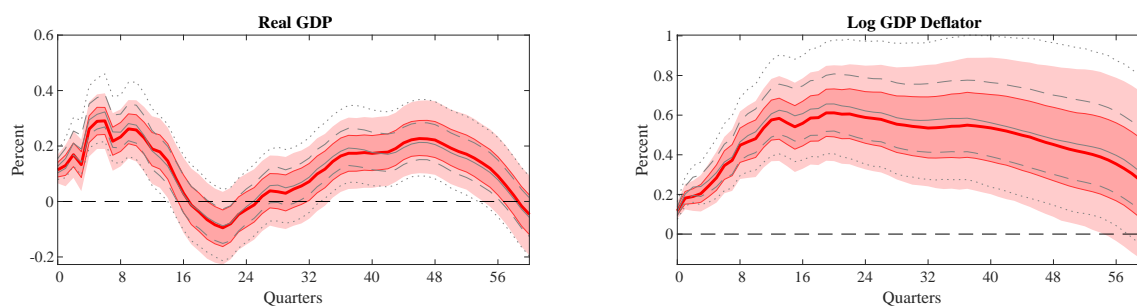


Note: The solid lines represent the median posterior response. The darker (lighter) shadow area represents the 68th (95th) HPD interval. All specifications use sixty lags of military spending news, government spending, GDP, GDP deflator, treasury bills, deficit to GDP ratio and debt to GDP ratio as in the baseline model. The prior uses a combination of “Minnesota”, “Sum of Coefficients”, and “Dummy Initial Observation” priors as described in the main text. The prior hyperparameters are $\mu = 10$, $\theta = 0.05$, with $\lambda = 0.56$.

L Excluding the Korean War

Another possible concern is that the exogenous variation in military spending news may be driven by some specific episodes such as the Korean war, thereby posing an external validity threat to the identification of government spending shocks via large war events and their estimated effects on output and prices. To ameliorate this concern, in this appendix we re-estimate our baseline VAR(60) but excluding the years of the Korean war, as in [Ramey and Zubairy \(2018\)](#). The findings in [Figure L.1](#) make clear that these make no material difference to our estimates.

Figure L.1: IMPULSE RESPONSE FUNCTIONS EXCLUDING THE KOREAN WAR



Note: The solid lines represent the median posterior response. The darker (lighter) shadow area represents the 68th (95th) 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.

M Impulse responses based on alternative GDP trend assumptions

Our baseline specification follows [Ramey and Zubairy \(2018\)](#) and therefore GDP enters the VAR (and the LP) as a ratio of the potential output estimated by these authors. But of course different assumption on the GDP trend may bear different implications on the low-frequency component and response of output and prices to any shock. Accordingly, in this appendix we consider four possible alternatives.

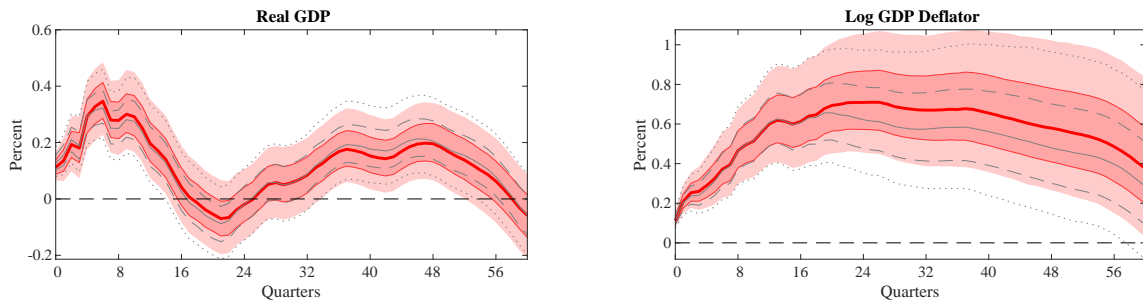
First, we consider the alternative measure of potential output proposed by [Ramey and Zubairy \(2018\)](#), which is constructed using a cubic trend for the early sample, fitted excluding Great Depression and spliced with the CBO potential GDP measure for the available years. Second, we use the first log difference of GDP in place of the ratio of GDP to potential output and also divide government spending by the values of GDP lagged by one quarter. Third, we estimate a specification in which real GDP per-capita enters the VAR in log-level, as it is the case for the GDP deflator. Finally, we adopt a version in which both real per-capita GDP and the GDP deflator enter the model in log differences.

The findings are reported in Figure [M.1](#), whose columns refer to output and prices respectively while the rows correspond to specifications using either [Ramey and Zubairy \(2018\)](#)'s alternative measure of potential (at the top) or the lagged values of GDP (second row) to normalize output and government spending, the log-level of GDP (third row) or both GDP and the GDP deflator in log differences (at the bottom). Red bands represent 68% and 90% credible sets for the posterior distributions of the impulse responses based on these alternative specifications. For the reader's convenience, we also report as grey lines the 68% and 90% confidence bands associated with the baseline specification in the main text.

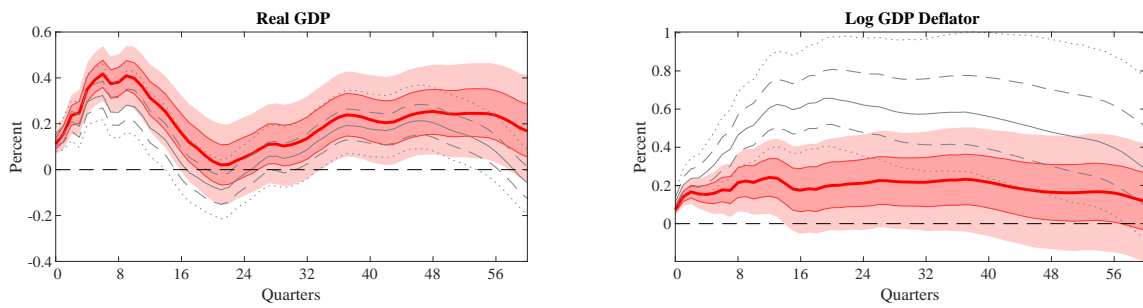
The very robust findings from this exercise is that none of these different trend assumptions on GDP overturn our main conclusions that government spending has large, significant and very persistent effects on both output and prices. In two out of the four specifications, the long-run effects of government spending on prices appear more muted and less persistent. It should be noted, however, that—even in these two specifications—the price response is still highly significant for more than five years after the shock, and therefore we that the latter finding as a robust inference one can draw on the low-frequency response of the price level to government spending.

Figure M.1: IMPULSE RESPONSE FUNCTIONS UNDER ALTERNATIVE TREND ASSUMPTIONS

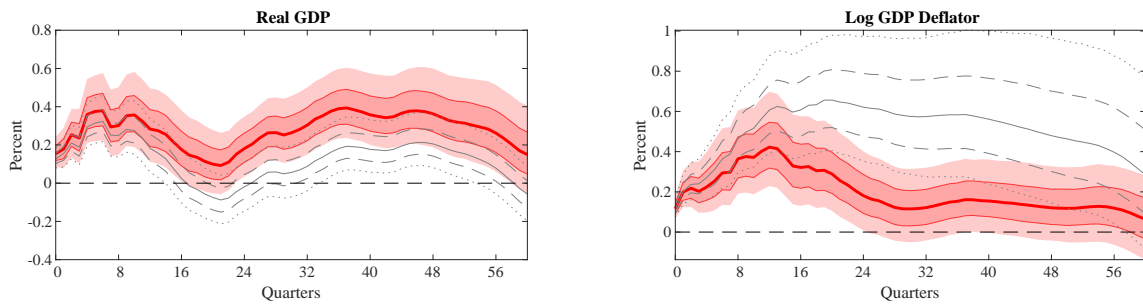
(a) Alternative Measure of Potential



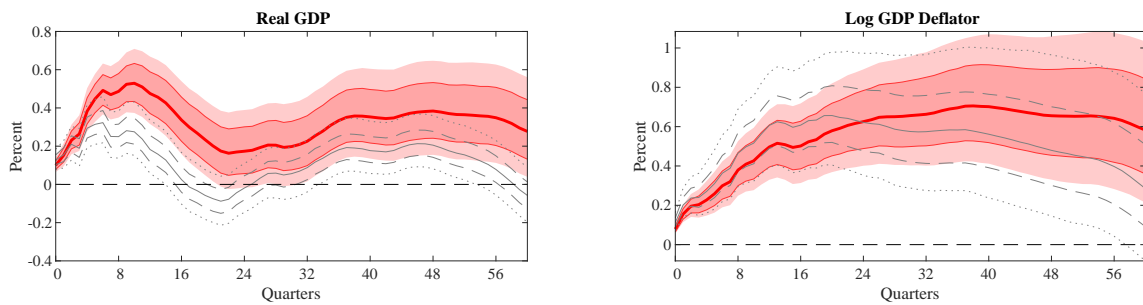
(b) Detrending by Previous Quarter GDP (stochastic trend)



(c) Real GDP per capita in log-levels (log-linear trend)



(d) Real GDP and GDP Deflator in Log Differences



Note: The solid lines represent the median posterior response. The darker (lighter) shadow area represents the 68th (95th) posterior credible set. 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 the first row, we use [Ramey and Zubairy's 2018](#) alternative measure of potential. In the second row, real GDP and government spending are both divided by GDP lagged by one quarter. In the third row, GDP enters the model in log levels and finally, in the fourth row, both GDP and the GDP deflator are specified in log-differences.

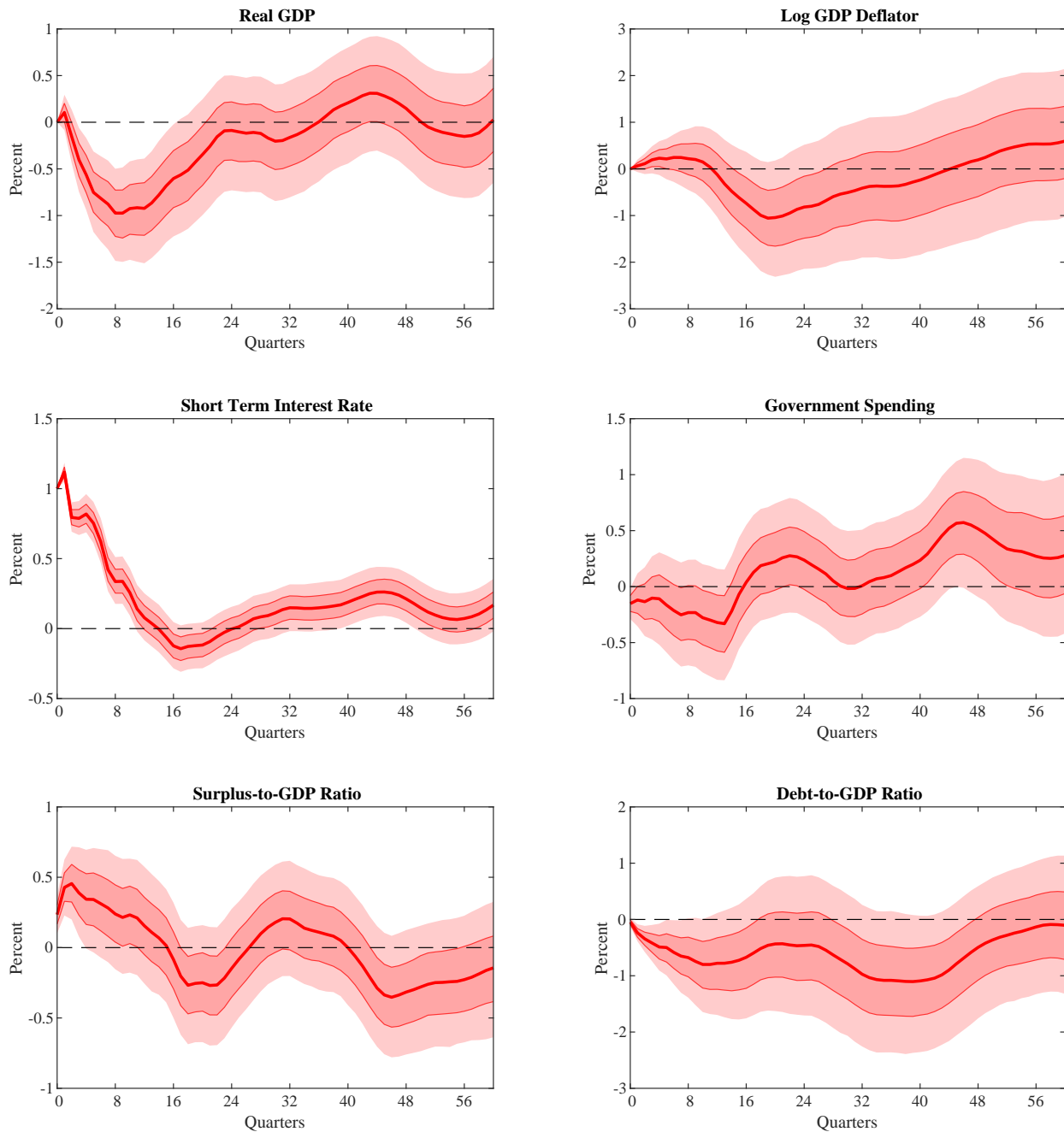
N Impulse Responses to Monetary Policy Shocks

We apply a Cholesky factorization to the estimates of a VAR(60) with the following variable order: real GDP per capita, GDP deflator, three months treasury bills, government spending, fiscal deficit to GDP ratio and public debt to GDP ratio.³ This amounts to the identification scheme for monetary policy shocks in [Christiano et al. \(2005\)](#), according to which the policy rate responds contemporaneously to movements in output and inflation whereas the latter respond only with a lag to movements in the interest rate.

The estimates in [Figure N.1](#) have all the hallmarks of the typical responses to a monetary policy shock. The effects on output are significant and short-lived, peaking around year 2. The effects on inflation are delayed, short-lived and peak four years after the shock. Using this identification scheme, we detect no low-frequency response of output and prices to a monetary policy shock. There is also some temporary contraction in government spending, which leads to a short-lived increase in the fiscal surplus and a reduction in public debt.

³For comparability with the other estimates in the paper, we have set the prior tightness, λ , to the value of 0.58, which is the optimal prior (according to the strategy in [Giannone et al., 2015](#)) that we have estimated for our baseline specification.

Figure N.1: IMPULSE RESPONSES TO A MONETARY POLICY SHOCK



Notes. The impulse responses are based on an estimated VAR with sixty lags of real per-capita GDP, GDP deflator, short-term interest rate, government spending, fiscal deficit to GDP ratio and government debt to GDP ratio. The short-term interest rate is ordered third in the Cholesky factorization, after real GDP and the GDP deflator. Output and government spending are expressed in percent of potential output as defined in [Ramey and Zubairy \(2018\)](#). The darker (lighter) shaded area represents the central 68% (90%) posterior band. The darker solid line stands for the median estimates. Results are based on 5000 posterior draws, discarding explosive roots.