Do Government Spending Multipliers Depend on the Sign of the Shock?

By Nadav Ben Zeev, Valerie A. Ramey and Sarah Zubairy*

Much recent attention has been devoted to estimating the size of government purchases multipliers. Part of that literature has explored whether multipliers are different during recessions or when monetary policy is constrained by the zero lower bound (e.g. Auerbach and Gorodnichenko (2012), Owyang, Ramey and Zubairy (2013), Ramey and Zubairy (2018)). An older literature explored asymmetry, i.e., whether declines in government spending have larger effects on economic activity than rises in government spending. For example, Hooker and Knetter (1997) found that military cutbacks had larger effects than military buildups on state economies and Davis, Loungani and Mahidhara (1997) found asymmetries in the effects of both oil shocks and government spending shocks on regional economies.

Recently, Barnichon, Debortoli and Matthes (2022) (BDM) offered evidence for asymmetric effects of government spending at the aggregate level. Using the Functional Approximations to Impulse Responses (FAIR) method of Barnichon and Matthes (2018) on aggregate U.S. data, BDM present evidence that negative shocks to government spending result in larger multipliers than positive shocks. Depending on the sample and identification method, they estimate multipliers between 0.3 and 0.8 for positive shocks to government spending and 1.4 for negative shocks. They also show robustness checks using nonlinear local projections.

Asymmetric government spending multipliers have first-order policy implications. If multipliers on rises in government spending are indeed less than multipliers on declines in government spending, then any government spending package that is not permanent will have a net negative effect on output. That is, the positive effect on output of an increase in government spending would be dominated by the negative effect of the unwinding of government spending. This result would imply two costs of a rise in government spending: the standard cost of eventually raising taxes to finance the spending plus the amplified negative effects of the wind-down of government spending. Thus, it is important to determine whether such asymmetries exist.

In this paper, we reexamine the evidence for asymmetric government spending multipliers in aggregate data. We first apply Ben Zeev's (2020) nonlinear diagnostic tests and find evidence of nonlinearities in the impulse response functions of both government spending and GDP. Since differences in impulse response functions do not necessarily translate into differences in multipliers, we explore the issue further by extending Ramey and Zubairy's (2018) framework to allow for asymmetric effects as a type of state dependence. While we find differences in the individual impulse response functions for positive versus negative shocks, the resulting multipliers do not differ by the sign of the shock.

We compare our results with those of BDM for our large historical sample and find that our local projection method produces more precise estimates of multipliers than their FAIR method, which is based on approximating the underlying impulse response functions. There is no evidence of difference in multipliers by sign in local projections. On balance, we conclude that the evidence for asymmetry is weak.

* Ben Zeev: Department of Economics, Ben-Gurion University of the Negev, P.O. Box 653, Beer-Sheva 84105, Israel, nadavbz@bgu.ac.il. Ramey: Department of Economics, 0508, UCSD, 9500 Gilman Drive, La Jolla, CA 92093, vramey@ucsd.edu. Zubairy: Department of Economics, 4228 TAMU, College Station, TX 77843, szubairy@tamu.edu. Ramey gratefully acknowledges financial support from NSF Grant No. 1658796.

I. Data

We use the data constructed by Ramey and Zubairy (2018), which consists of quarterly data from 1889 to 2015 on GDP, government purchases, and a narrative military news series. The military news series, which is constructed from narrative evidence, consists of changes in the expected present discounted value of the path of government purchases. All three variables are first deflated by the GDP price deflator and then divided by an estimate of potential GDP, based on a sixth degree polynomial trend fit over the sample excluding the Great Depression and WWII (1930 - 1946).

II. A Case Study of Two World Wars

To motivate our skepticism of asymmetric effects on multipliers, we first review events during two influential episodes in the U.S. historical data. Figure 1 shows the behavior of real government purchases and real GDP, both divided by potential GDP.

In each case, government spending rose when the U.S. became involved in the war. At the end of the war, real government spending returned to its pre-war fraction of potential GDP in less than a year.

GDP rose and fell along with government spending. If anything, GDP rose more with the rise in government spending than it fell with the decline in government spending at the end of the war. In both cases, the recession at the end of the war was shallow and brief.[1] Thus, there is no evidence in the raw data suggesting that declines in government spending have greater effects than rises in government spending.

III. Diagnostic Tests for Nonlinearities

We conduct some initial diagnostic tests for nonlinearity using Ben Zeev's (2020) polynomial test, augmented with Forni et al.'s (2022) method for differentiating nonlinearities due to sign versus size. Impulse response functions (IRFs) are estimated using local projections on a set of regressions for each horizon h, from 0 to 20 quarters, as follows:

$$news_t = \delta(L) z_{t-1} + \eta_t \tag{1}$$

$$x_{i,t+h} = \alpha_{i,h}\hat{\eta}_t + \theta_{i,h} f(\hat{\eta}_t) + \psi_{i,h}(L) z'_{t-1} + \zeta_{i,t+h}, \quad \textit{for } i = g,y \tag{2}$$

In Equation 1, *news* is the military news variable and z consists of a constant term plus four lags of news, government spending, and GDP, all transformed as described in the data section. Equation 2 represents two additional equations where x is government spending (g) in one equation and GDP (y) in the other. $f(\hat{\eta}_t)$ is the nonlinear term in the shock.[2] The z' is z augmented with lags of the nonlinear shock term.

We first test the null hypothesis $\theta = 0$ against the quadratic alternative $f(\hat{\eta}_t) = \hat{\eta}_t^2$. We reject the null hypothesis at the 5% (10%) level for 13 (18) of 20 horizons for government spending and 8 (9) for GDP, suggesting nonlinearities in the IRFs for both variables. The online appendix shows the estimated IRFs.

However, the quadratic term could be picking up sign or size effects. To distinguish asymmetric effects from size nonlinearities, we conduct a test using Forni et al.'s (2022) absolute value term, $f(\hat{\eta}_t) = |\hat{\eta}_t|$, which captures only asymmetry. We find that this term is also significant at almost all horizons for both variables. Following Forni et al. (2022), we run a horse race between the quadratic and absolute value terms by including both terms in the model. The correlation between the two nonlinear terms is 0.9, so multicollinearity results in neither being individually significant in 28 of 42 cases. The absolute value term is significant in all 14 of the others cases, but the quadratic term is significant in only 2 cases. These results suggest that sign not size is the main source of nonlinearity.

However, differences in IRFs do not imply differences in multipliers since the mul-

[1] In the case of WWI, a deeper recession followed in 1920-21 but is usually attributed to severe monetary tightening.

[2] We use the innovation in the news equation as the shock because the polynomial method requires a mean-zero shock. In the next section, we use news itself.

tiplier is based on the ratio of the IRFs. To investigate whether multipliers differ by the sign of the shock, the next section develops a framework for estimating both IRFs and multipliers when there are asymmetries.

IV. State-Dependent Local Projections

A. Econometric Model

Asymmetry can easily be modeled by redefining the state in Ramey and Zubairy's (2018) (RZ) framework. A key advantage of the RZ framework is the equivalence of the three-step and one-step estimates of multipliers, which facilitates estimation of standard errors on multipliers and tests of equality of multipliers.

The impulse response functions are estimated with a set of regressions for each horizon h using the following model:

(3)
$$x_{i,t+h} = I_t^+ \left[\beta_{i,h}^+ news_t + \phi_{i,h}^+(L) z_{t-1}\right] + I_t^- \left[\beta_{i,h}^- news_t + \phi_{i,h}^-(L) z_{t-1}\right] + \varepsilon_{i,t+h}, \quad \textit{for } i = g,y \textit{ and } h = 0,\ 1,\ \ldots,\ H$$

Here x is either government spending (g) or GDP (y) and I^+ is a dummy variable for $news_t > 0$ and I^- is its complement. z consists of a constant term and four lags of government spending, GDP, and news. All the coefficients of the model are allowed to differ according to whether the contemporaneous shock is positive or negative; this flexibility is a necessary condition for the cumulative multiplier computed from the estimated IRFs to be equivalent to the one-step IV multiplier defined below.

The cumulative multiplier through horizon h is the ratio of the integral under the GDP IRF to the integral under the government spending IRF, i.e., the multiplier for positive shocks is $m^+ = \left(\sum_{j=0}^{h} \beta_{y,h}^+\right) / \left(\sum_{j=0}^{h} \beta_{g,h}^+\right)$ and similarly for negative shocks. This three-step method for computing multipliers produces point estimates, but obtaining standard errors and doing tests for the equality of the two multipliers is cumbersome. An easier method is the one-step local projection-instrumental variables (LP-IV) method introduced by Ramey and Zubairy (2018). This procedure involves IV estimation of a regression of the cumulative sum of GDP on the cumulative sum of government spending using the shocks as instruments. In particular, we estimate

(4)
$$\sum_{j=0}^{h} y_{t+j} = m_h^+ \left(I_t^+ \sum_{j=0}^{h} g_{t+j} \right) + m_h^- \left(I_t^- \sum_{j=0}^{h} g_{t+j} \right) + I_t^+ [\gamma_h^+(L) z_{t-1}] + I_t^- [\gamma_h^-(L) z_{t-1}] + \omega_{t+h}, \quad \textit{for } h = 0,\ 1,\ \ldots, H$$

using $news_t^+$ and $news_t^-$ as instruments for the terms in parenthesis. The cumulative multiplier through horizon h is the coefficient m_h^+ for positive shocks and m_h^- for negative shocks.[3]

B. Results

Figure 2 shows the IRFs in response to positive and negative shocks. The estimates imply that a unit-magnitude negative shock leads to much larger responses of both government spending and GDP, though the estimates are less precise for negative shocks than positive shocks. However, Figure 3 shows that these differences do not translate into significant differences in multipliers since both the numerator and denominator of the multiplier increase roughly proportionally for negative shocks. We fail to reject equality of the multipliers for all horizons other than the first couple quarters, which display the typical pattern that GDP responds more quickly than government purchases to a news shock. The rise in GDP appears as inventory investment in the short run and in government purchases only after the goods have been delivered (Briganti and Sellemi (2022)). This accounting feature explains why Cholesky decompositions on government spending get

[3]We use Newey-West corrections of the standard errors in IRF and multiplier regressions rather than lag-augmented regressions because of apparent additional sources of serial correlation.

the timing of the shock wrong.

In sum, the findings are similar in spirit to RZ's finding for slack states: news shocks during both slack and negative shock states generate bigger changes in government spending. However, GDP rises proportionally, so there is no difference in multipliers. The online appendix shows that these results are robust to many variations.

V. Comparison to BDM Results

Our results contrast with Barnichon, Debortoli and Matthes's (2022) (BDM) results, so we investigated possible sources for the differences. We limited our comparisons to specifications that use RZ's data, historical sample, and narrative military news shock. Details of the results are reported in our online appendix.

BDM use functional approximations to impulse responses (FAIR) to estimate impulse responses and then construct multipliers from those. As BDM note, the FAIR method likely induces bias, but they advocate its use based on efficiency gains.

The left panel of Figure 4 shows our replication of BDM's FAIR results. The FAIR estimates are less precise than the LP-IV multipliers shown in Figure 3. However, part of the difference is how negative and positive shocks are defined. 78 percent of the military news observations are zeroes, so it matters how they are grouped. Based on the paucity of strictly negative shocks (only 6.5 percent of the sample) and the low F-statistics for the first stage for strictly negative shocks, we grouped the zeroes with the negative shocks in our baseline model. In contrast, BDM grouped the zeroes with the positive shocks. The right panel of Figure 4 shows alternative LP-IV model estimates that follow BDM in grouping the zeroes with the positive values. A comparison of this graph with our baseline model shows that BDM's grouping lowers the precision. Neverthless, these alternative LP-IV estimates are still more precise than the FAIR estimates on average.

As shown by Plagborg-Møller and Wolf (2021), VARs and LPs estimate the same impulse response functions in population, so with our controls similar to a VAR equation, we likely have efficiency gains over the MA representation of BDM because we have such a large sample (500 observations).

BDM check for possible bias in their FAIR estimates by estimating a local projections model, and again find differences in point estimates of multipliers by sign. Our examination of their replication programs, however, revealed some issues in implementation, which are detailed in the online appendix. Furthermore, our statistical tests were unable to reject equality of multipliers at any horizon for any significance level below 30 percent.

In sum, the estimates from our state-dependent local projections are subject to less bias and we find that they are more precise than the FAIR estimates in this setting. Our estimates suggest no asymmetry in government spending multipliers.

REFERENCES

Auerbach, Alan J, and Yuriy Gorodnichenko. 2012. "Measuring the output responses to fiscal policy." *American Economic Journal: Economic Policy*, 4(2): 1–27.

Barnichon, Regis, and Christian Matthes. 2018. "Functional approximation of impulse responses." *Journal of Monetary Economics*, 99: 41–55.

Barnichon, Regis, Davide Debortoli, and Christian Matthes. 2022. "Understanding the size of the government spending multiplier: It's in the sign." *The Review of Economic Studies*, 89(1): 87–117.

Ben Zeev, Nadav. 2020. "Identification of Sign Dependency of Impulse Responses." *Ben-Gurion University of Negev working paper.*

Briganti, Edoardo, and Victor Sellemi. 2022. "Why Does GDP Move Before G? It's All in the Measurement." *UCSD working paper.*

Davis, Steven J, Prakash Loungani, and Ramamohan Mahidhara. 1997. "Regional labor fluctuations: oil shocks, military spending, and other driving forces." *Board of Governors of the Federal Reserve System International Finance Disc. Papers*, 578.

Forni, Mario, Luca Gambetti, Nicolò Maffei-Faccioli, and Luca Sala. 2022. "The Nonlinear Transmission of Financial Shocks: Some Evidence." *Norges Bank working paper*.

Hooker, Mark A, and Michael M Knetter. 1997. "The effects of military spending on economic activity: Evidence from state procurement spending." *Journal of Money, Credit, and Banking*, 400–421.

Owyang, Michael T, Valerie A Ramey, and Sarah Zubairy. 2013. "Are government spending multipliers greater during periods of slack? Evidence from twentieth-century historical data." *American Economic Review*, 103(3): 129–34.

Plagborg-Møller, Mikkel, and Christian K. Wolf. 2021. "Local Projections and VARs Estimate the Same Impulse Responses." *Econometrica*, 89(2): 955–980.

Ramey, Valerie A., and Sarah Zubairy. 2018. "Government spending multipliers in good times and in bad: evidence from US historical data." *Journal of Political Economy*, 126(2): 850–901.

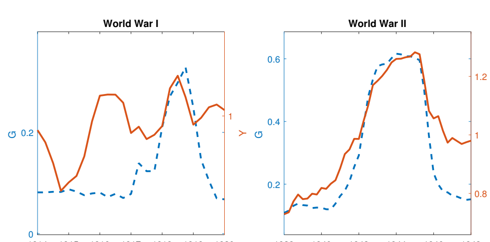

Figure 1. Case Study of Two World Wars.

Note: Real government purchases (dashed) and GDP (solid) divided by potential GDP. Source: Ramey and Zubairy (2018) data.

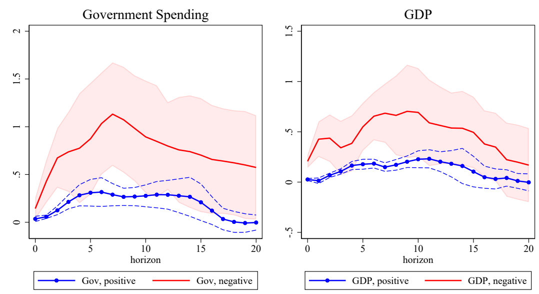

Figure 2. Responses of G and Y to Military News

Note: 95% confidence bands.

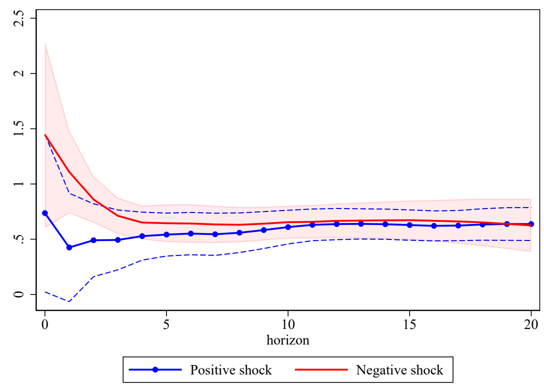

Figure 3. LP-IV Multipliers by Sign

Note: 95% confidence bands.

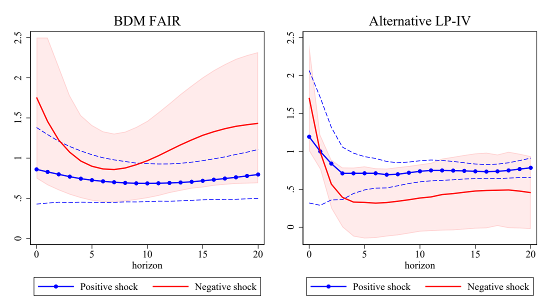

Figure 4. BDM FAIR vs. Alternative LP-IV Multipliers

Note: 95% bands. Both models include zeroes with the positive shock.