Online Appendix for Do Government Spending Multipliers Depend on the Sign of the Shock?

Nadav Ben Zeev
Ben-Gurion University of the Negev

Valerie A. Ramey
University of California, San Diego, NBER, and CEPR

Sarah Zubairy
Texas A&M University and NBER

March 6, 2023

Valerie Ramey gratefully acknowledges financial support from National Science Foundation Grant No. 1658796.

Contents

List of Tables

List of Figures

1 Extended Local Projection Results and Robustness Checks

In this section, we provide more complete results for our baseline model, including first-stage F-statistics, and also check the robustness of our results to alternative specifications. We find no differences in multipliers by sign of shock in these alternative specifications. In addition, we compare the precision of our results to those from the Barnichon et al. (2022) FAIR method.

1.1 Baseline Model and Results

As discussed in the paper, in our baseline model the impulse response functions are estimated with a set of regressions for each horizon h using the following model:

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

Here x is either government spending (g) or GDP (y), both normalized by potential GDP as in Ramey and Zubairy (2018). z consists of a constant term and four lags of government spending, GDP, and news. We define I^+ as a dummy variable for $news_t > 0$ and I^- is its complement.

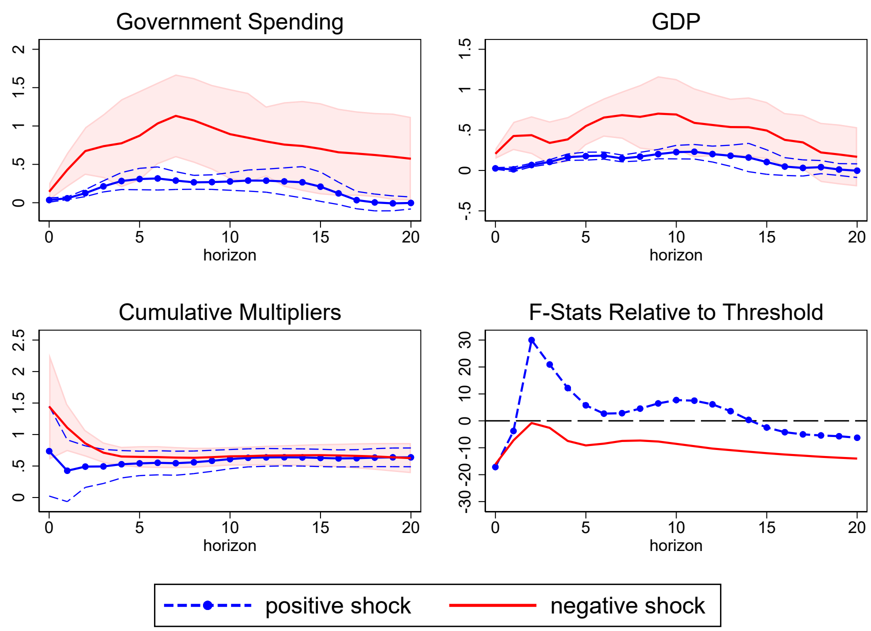
To estimate multipliers, we employ a one-step local projection-instrumental variables (LP-IV) method introduced by Ramey and Zubairy (2018) to compute cumulative multipliers. This procedure involves IV estimation of a regression of the cumulative sum of GDP on the cumulative sum of government spending using the shocks as instruments. In particular, we estimate

$$\sum_{j=0}^{h} y_{t+j} = m_h^+\left(I_t^+\sum_{j=0}^{h} g_{t+j}\right) + m_h^-\left(I_t^-\sum_{j=0}^{h} g_{t+j}\right) + I_t^+[\gamma_h^+(L)z_{t-1}] + I_t^-[\gamma_h^-(L)z_{t-1}] + \omega_{t+h}, \quad \text{for } h = 0, 1, ..., \text{H} \tag{2}$$

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.

Figure 1 shows the impulse response functions (IRFs) and multipliers from the text, as well as the first-stage F-statistics. As discussed in the text, the IRFs for both govern-

Figure 1. Baseline Results



Note: Responses to military news, baseline results reported in the paper. 95% confidence bands.

ment spending and GDP are greater after a negative shock than a positive shock. However, those differences in IRFs do not translate into differences in multipliers. Other than the first few quarters, which are affected by output jumping in anticipation of the rise in government spending, there is no quantitative or statistical difference in multipliers by sign of the shock.[1]

Figure 1 also shows the first-stage F-statistics for the LP-IV regression that estimates cumulative multipliers. Because the military news variable is based on changes in defense spending due to political events, it should be exogenous to the economy. Since there is inherent serial correlation based on using the Jordà method, we use the Montiel Olea and Pflueger (2013) effective F-statistics and thresholds, following the approach taken in Ramey and Zubairy (2018). The lower right panel shows the *difference* between the first-stage effective F-statistics and the Montiel Olea and Pflueger (2013)

1. The behavior during the first few horizons is due to GDP jumping before government spending does. Briganti and Sellemi (2022) document that this anticipation effect is due to national income accounting conventions. In particular, production occurs as soon as the government contract is awarded and is counted in inventory investment. The production does not show up in government spending until several quarters later when the final goods are shipped to the government and the government pays the company.

thresholds.[2] A value above 0 means that the effective F-statistic exceeds the threshold. The F-statistics is above the threshold for most horizons for the positive shock, but is just below for the negative shock.

1.2 Alternative Definitions of State

In our baseline specification, we define I^+ as a dummy variable for $news_t > 0$ and I^- is its complement. We also use military news itself to define the state. In this section, we explore two alternatives to this baseline.

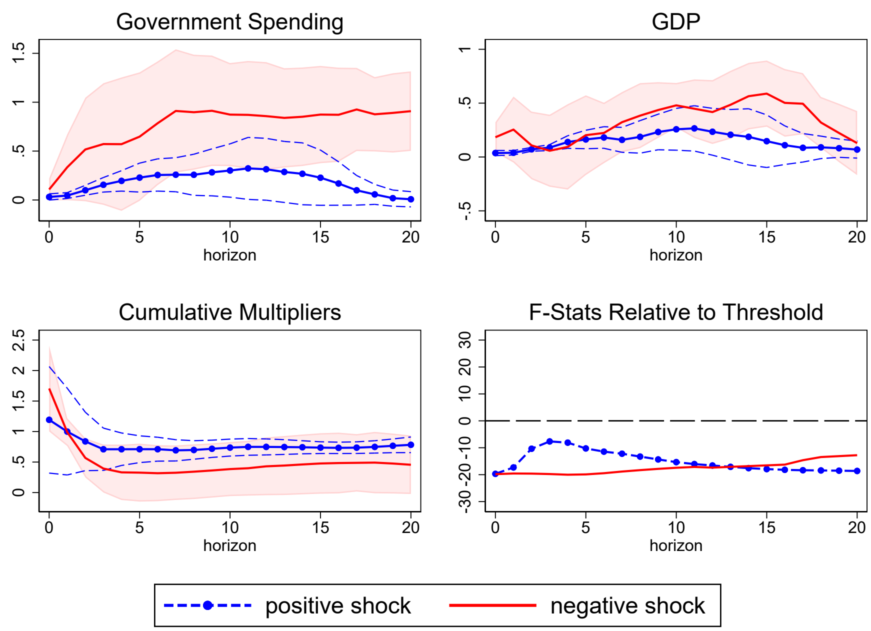
The first alternative concerns the choice of how to group the zeroes of news. In the baseline specification we group the zeroes with negative news. This decision is driven by the fact that we have only 33 negative news shocks (6.5% of the observations) and we have 75 positive shocks (14.8 % of the observations); the remaining 397 observations are zeros. This choice improves the first-stage F-statistics in the multiplier estimation.

Here we consider the alternative specification used by Barnichon et al. (2022), which in our model is equivalent to defining I^+ as a dummy variable with a weak inequality, i.e., $news_t \geq 0$. The impulse responses and multipliers for this alternative are shown in Figure 2. Similar to our baseline results, the IRFs to the negative shock are larger, but they do not result in multipliers that are larger. In fact, switching to BDM's definition of positive and negative results in point estimates that imply slightly higher multipliers for the positive shock, though they are not statistically different. This alternative specification also results in larger confidence bands, particularly for negative shocks, and lower first-stage F-statistics.

More generally, our explorations show that both IRFs and multipliers tend to be estimated more precisely when we use our baseline grouping of zero values of news. Table 1 shows the average standard errors of the estimates of multipliers over a horizon of 20 quarters for the Barnichon et al. (2022) (BDM) FAIR model, our baseline LP-IV, and alternative LP-IV models. Our baseline specification for the LP-IV has the lowest average standard errors for the multiplier estimates. Even with the alternative definition of state that matches BDM, our LP-IV has lower average standard errors than BDM's FAIR method, even though the latter involves estimating a smaller number of parameters because it approximates the underlying impulse response functions.

2. We use the threshold for the 5 percent critical value for testing the null hypothesis that the TSLS bias exceeds 10 percent of the OLS bias. For one instrument, this threshold is always 23.1. The threshold is 19.7 percent for the ten percent critical value.

Figure 2. Results Using Alternative Grouping of Zero Values of News



Note: Responses to military news, when we define I^+ as the dummy variable for $news_t \geq 0$ to define positive shocks as states. 95% confidence bands.

Table 1. Average Standard Errors of Multiplier Estimates (Horizon 0 to 20)

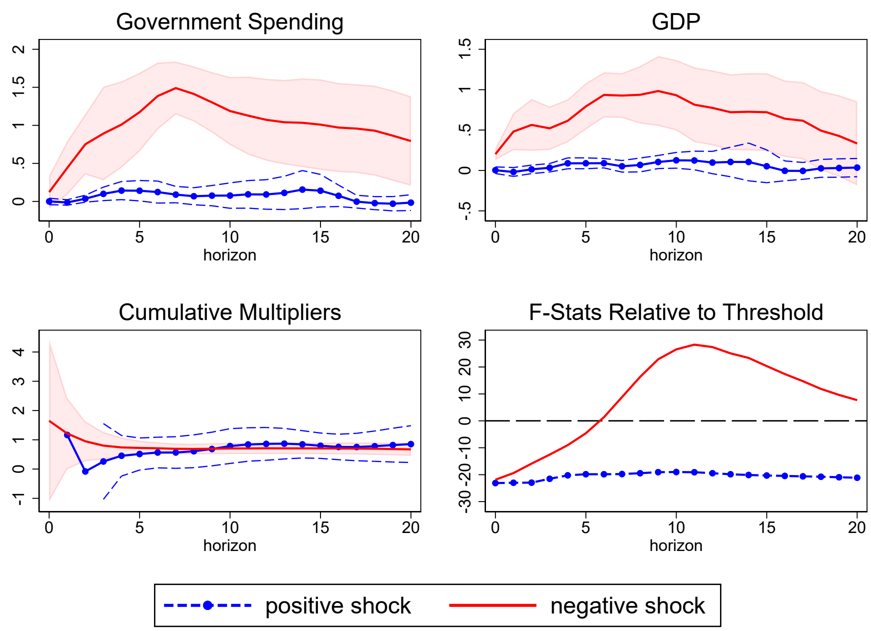
Specification	Negative shocks	Positive shocks
Baseline LP-IV ($I^+ = news > 0$)	0.11	0.11
Alternative LP-IV ($I^+ = news \geq 0$)	0.23	0.12
BDM FAIR ($I^+ = news \geq 0$)	0.33	0.14

The second alternative we consider is a specification where we use the innovation to $news_t$, η_t, to instrument for the government spending shock instead of $news_t$ itself, estimated from the following equation:

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

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 of the paper. In the paper, we used the mean-zero innovation to news in our diagnostic tests for non-linearity in the paper. In this specification, grouping of zeroes is not an issue in defining the state since no innovation is identically zero.

Figure 3. Results with the Innovation to News as the Shock



Note: Responses to military news, when we consider $\hat{\eta}_t$, residual from Equation (3), as instrument for government spending shocks. 95% confidence bands.

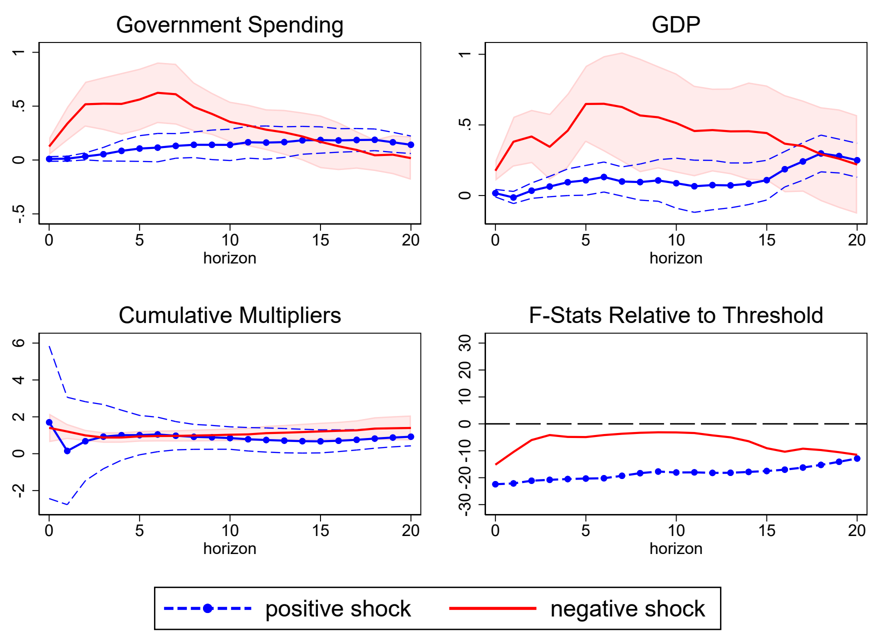
The results are shown in Figure 3. Both the impulse responses and multipliers look similar to the baseline. The only difference is that the confidence bands are very large for the first few horizons of the positive shock multiplier — we have set them to missing so that the scale of the graph is more informative. The F-statistics change noticeably. Here, the F-statistics for the negative shock rise above the threshold by six quarters, but the positive shock is always below the threshold.

1.3 Results with Barnichon, Debortoli and Matthes' (2022) Additional Controls

Our baseline model follows Ramey and Zubairy (2018) by including four lags of government spending, GDP, and military news as controls. On the other hand, Barnichon et al. (2022) also include taxes and quartic trend variables in their LP specification and use 8 lags rather than 4. In order to assess the effects of these additional variables, we increase the lags in our model from 4 to 8 and also include quartic trends and lags of taxes.

The results for the impulse response functions and the multipliers are shown in Figure 4. The multipliers are neither quantitatively or statistically different for positive versus negative shocks at any horizon; the p-value for equality never falls below 0.16 according to our tests of equality (not shown). Also, as shown in the bottom right panel of the figure, adding additional lags and controls leads to instrument relevance issues for both positive and negative shocks.[3]

Figure 4. Our Baseline Model with Barnichon et al. Controls



Note: Responses of G and Y to military news, using Barnichon et al.'s (2022) choice of 8 lags, taxes as controls and quartic trend. 95% confidence bands.

1.4 Hall-Barro-Redlick Transformation

In order to estimate multipliers, one should not use logarithms of variables since they lead to biased multiplier estimates (Owyang et al. (2013)). However, since GDP and government spending have exponential trends, some kind of transformation is required. Our baseline specification and Barnichon et al.'s (2022) specifications follow

3. These results are based on grouping zero news with negative news, as in our baseline model. If we instead group zero news with positive news as in BDM, and also include their 8 lags, taxes as additional controls and quartic trends, we run into singular matrices. The reason is that there are so few strictly negative shocks (only 6.5 % of observations) so the large number of controls saturate the data.

Ramey and Zubairy (2018) in using a Gordon and Krenn (2010) transformation, which normalizes GDP, government spending, and military news by a measure of potential GDP. Our baseline measure of potential real GDP is from a regression on a sixth-degree polynomial trend that omits the Great Depression and World War II. To make sure our results are not being affected by the normalization by potential GDP (which implicitly does not allow government spending to affect potential GDP) and to avoid any use of deterministic trends, we re-estimate our model using the transformation used by Hall (2009) and Barro and Redlick (2011). This transformation uses cumulative differences of real government spending and GDP, and divides both by lagged real GDP. We also divide the military news variable by lagged nominal GDP (since military news is nominal). In particular, the local projections for estimating the IRFs are:

$$\frac{x_{i,t+h} - x_{i,t-1}}{y_{t-1}} = I_t^+ \left[\beta_{i,h}^+ news_t + \phi_{i,h}^+(L) z_{t-1}\right] + I_t^- \left[\beta_{i,h}^- news_t + \phi_{i,h}^-(L) z_{t-1}\right] + \varepsilon_{i,t+h}, \quad \textit{for } i = g, y \textit{ and } h = 0, 1, \ldots, H \tag{4}$$

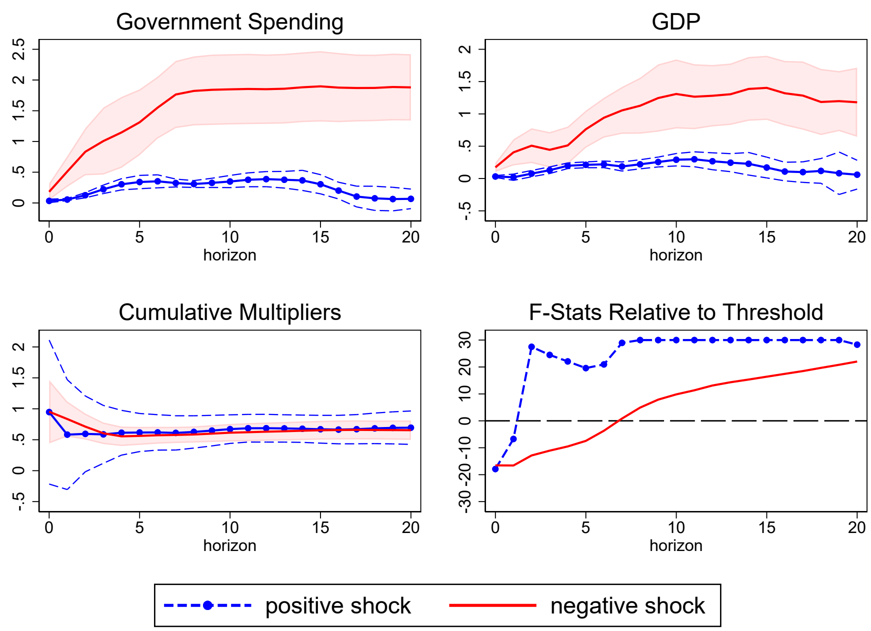
Here the x is either real government spending (g) or real GDP (y), with no potential GDP normalization as in the baseline. z consists of a constant term and lagged values of military news (divided by lagged nominal GDP), as well as lags of first differences of real government spending and GDP, both divided by lagged real GDP. The equation for the one-step LP-IV estimate of the cumulative multipliers is:

$$\sum_{j=0}^{h} \frac{y_{t+j} - y_{t-1}}{y_{t-1}} = m_h^+ \left(I_t^+ \sum_{j=0}^{h} \frac{g_{t+j} - g_{t-1}}{y_{t-1}} \right) + m_h^- \left(I_t^- \sum_{j=0}^{h} \frac{g_{t+j} - g_{t-1}}{y_{t-1}} \right) + I_t^+[\gamma_h^+(L) z_{t-1}] + I_t^-[\gamma_h^-(L) z_{t-1}] + \omega_{t+h}, \quad \textit{for } h = 0, 1, \ldots, H \tag{5}$$

using $news_t^+$ and $news_t^-$ as instruments for the terms in parenthesis.

Figure 5 shows the results. The impulse responses of government spending and GDP to the negative shocks suggest very persistent effects on both series, whereas the positive shocks display more mean reversion. However, both imply the same values of cumulative multipliers by sign of shock, similar to our baseline. Interestingly, the first-

Figure 5. Results using Hall-Barro-Redlick Transformation



Note: Responses to military news using the Hall-Barro-Redlick transformation of variables in our baseline specification.

stage F-statistics are significantly higher for both positive and negative shocks than for our baseline specification.

2 Comments on Barnichon, Debortoli and Matthes' (2022) Local Projections Specification

Barnichon et al. (2022) (BDM) recognize that their FAIR method, which uses functional approximations, likely induces bias. Thus, they conduct a robustness check using the Ramey and Zubairy (2018) (RZ) local projections model, where they also find differences in point estimates of multipliers. Focusing on their analysis that uses the historical data, we examined their replication programs and comment on several details of their implementation.

First, BDM do not allow all the coefficients in their nonlinear LP model to vary with the sign of the shock, resulting in a breakdown in the equivalence of the variations on the one-step LP-IV multiplier estimation method and the three-step method.[4] In Ramey and Zubairy's (2018) and Ben Zeev, Ramey and Zubairy (2023) (BRZ)'s state-dependent models, which allow all coefficients to change with the state, the estimate of the multiplier will be equivalent whether one uses the three-step method or one-step LP-IV method. Because BDM's model does not have this equivalence, they are unable to conduct statistical tests of the equality of multipliers by sign of the shock.

Second, the LP model estimated by BDM differs from both their FAIR model and the RZ baseline model, as we show in Table 2. BDM include taxes and a quartic trend in their LP model, but not in their FAIR model with narrative identification.[5] They include the quartic trend despite following RZ in normalizing all their variables by a sixth-degree polynomial trend estimate of potential GDP. The BDM LP specification also deviates from the RZ specification by using many more lags (8 rather than 4). The combination of extra variables, extra lags results, and a quartic trend results in very imprecise estimates in BDM's LP model. As our robustness checks in the last section show, adding their controls to our baseline model results in multipliers that are quantitatively and statistically indistinguishable by sign. In additional investigations, we found that the quartic trend in particular can lead to odd behavior at the longer horizons on which BDM focus (i.e. quarter 20). Our investigations suggest that their process of detrending the ratio of government spending to potential GDP deforms the series and produces estimates of multipliers for negative shocks that rise after 10 quarters (see the lower right panel of Figure 4 of their paper). As shown in Section 1.3, even when we add their additional controls to our one-step LP-IV multiplier model, we do not find any evidence of statistically significantly larger multipliers for negative shocks relative to positive shocks.

4. The three-step method of computing multipliers estimates the IRFs of government spending and GDP separately via OLS, calculates the integrals under the respective IRFs, and forms their ratios.

5. They include taxes only in their recursively identified model because they follow Blanchard and Perotti's Cholesky decomposition method.

Table 2. Discrepancies across FAIR and LP specifications in Barnichon, Debortoli and Matthes (2022), and comparison with Ramey and Zubairy (2018).

	BDM - FAIR	**BDM - LP**	**RZ - LP**
Lags	N/A	8	4
Trends	None	Quartic trends	None
Taxes	None	Yes	None
Sample (text)	1890-2014	1890-2014	1890-2015
Sample (programs)	1901-2015	1890-2015	1890-2015

3 Barnichon, Debortoli and Matthes (2022) FAIR results employing narrative identification: Robustness

The baseline results for Barnichon et al. (2022) (BDM) showing evidence of larger multipliers as a result of negative shocks than positive shocks are based on their FAIR approach. In this section, we revisit those findings. We first point out that they use a slightly shorter sample than the original RZ data set, which is in contrast to what is mentioned in their text. We then document the impact of using the full RZ sample on their main findings. Next, we test the robustness of their results to the choice of K, which is an important parameter in their moving average approximation of the impulse responses for their FAIR methodology. We show that both these departures weaken the statistical significance of their main finding of larger multipliers for negative shocks than positive shocks.

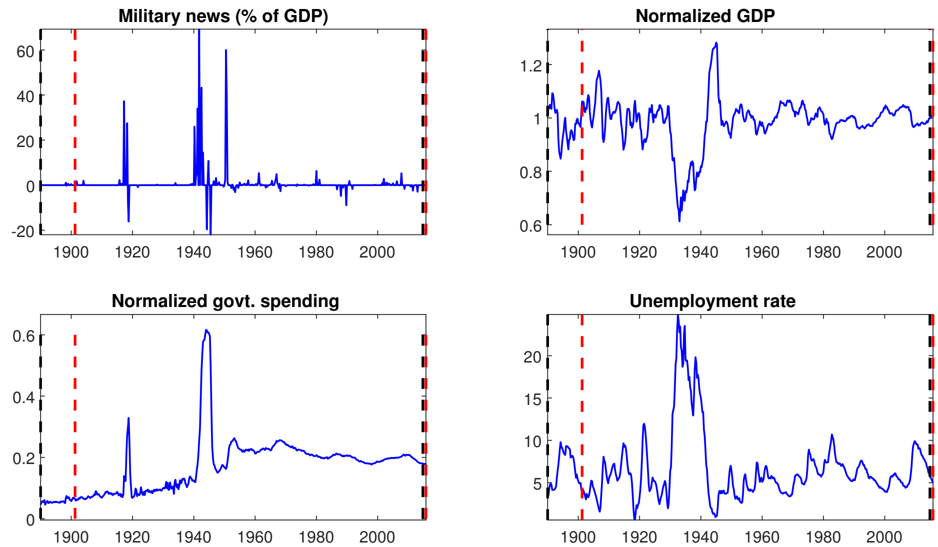
3.1 Sample choice

The Functional Approximations to Impulse Responses (FAIR) results employing the narrative identification scheme shown in Figure 3 of the BDM paper claims that it uses the Ramey and Zubairy (2018) data set, and employs data from 1890-2014. To be precise, Ramey and Zubairy (2018) construct and use the data from 1890q1-2015q4. However, the codes used to generate Figure 3 of Barnichon et al. (2022) employs data from 1901q2-2015q4.

If we keep all details of the analysis behind Figure 3 of the BDM paper the same, and employ the entire RZ sample as they claim, the results are slightly weaker. Notably, the evidence of the posterior probability that the multiplier for a contractionary shock is larger than an expansionary shock is above 0.95 from 14 quarters onwards until 30 quarters with the sample they use, that spans 1901q2-2015q4. With the full sample this posterior probability is above 0.95 only for quarters between quarter 17 and 24.

The posterior probability that the multiplier for a contractionary shock is larger than an expansionary shock at horizon 20 is 0.9758 for their sample, and is 0.9514 for the full RZ sample.

Figure 6. Data used in the analysis in Section 3.3. of Barnichon et al. (2022)



Note: The figure shows the sample employed by Ramey and Zubairy (2018), which spans 1890q1-2015q4. BDM claim that Figure 3 in their paper is generated using the sample period given within the black dashed bars, 1890q1-2014q4. The sample employed by BDM in the codes that generates Figure 3 is given within the red dashed bars, and spans 1901q2-2015q4.)

3.2 Estimation Methodology: choice of MA lags

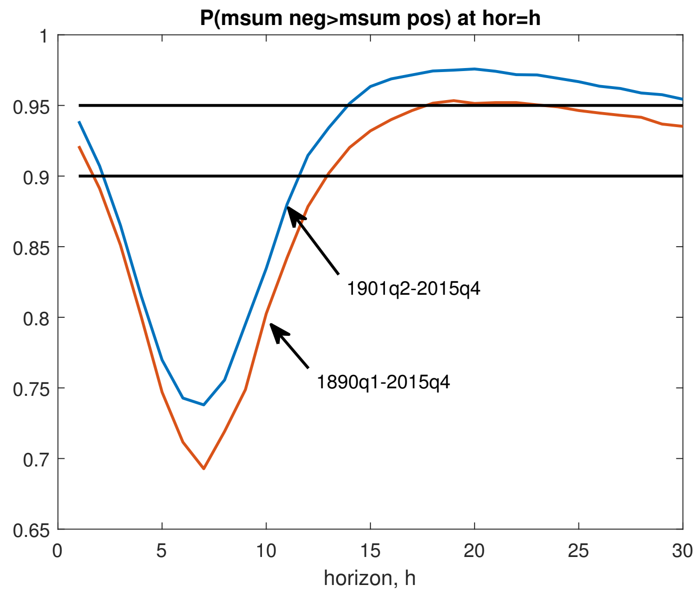
BDM employ the FAIR estimation methodology for their baseline results. One of the major benefits of this approach touted in the paper is that it reduces the number of parameters to be estimated substantially. In order to see this consider a simple linear case, where y_t is a vector of stationary macroeconomic variables, and it can be written as a structural moving-average (MA) model

$$y_t = \sum_{k=0}^{K} \Psi_k \epsilon_{t-k} \tag{6}$$

where ϵ_t is a vector of i.i.d structural shocks and K is the number of lags, which can be finite or infinite.

Denote $\psi(k)$ as an element of matrix Ψ_k, so that $\psi(k)$ is the impulse response function ψ at horizon k. The FAIR approach provides the following functional approxima-

Figure 7. FAIR results employing narrative identification scheme: sample choice



Note: The figure shows the posterior probability that the cumulative multiplier for a contractionary shock is larger than an expansionary shock at a given horizon, h. The two different lines corresponds to the sample employed in Barnichon et al. (2022) code, which spans 1901q2-2015q4 (blue line), and when we employ the full Ramey and Zubairy (2018) sample to conduct the same exercise, which spans 1890q1-2015q4.)

tion:

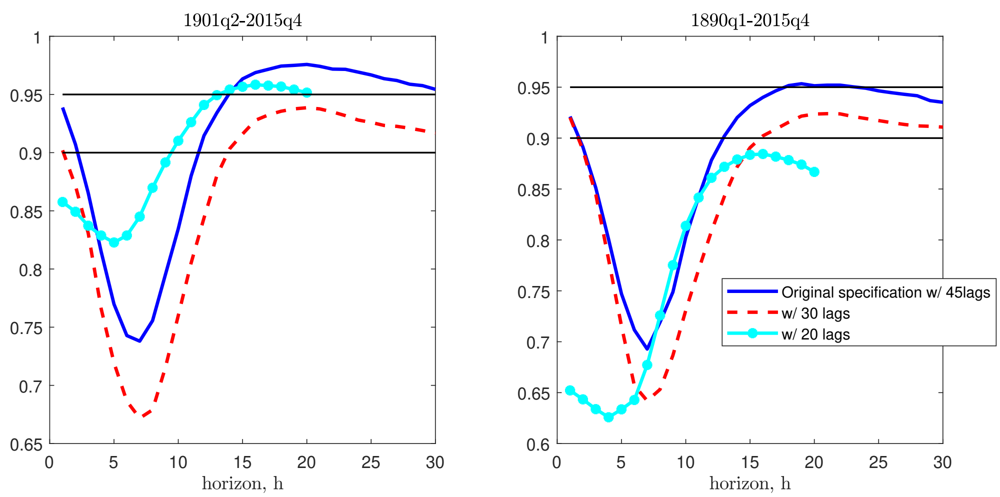
$$\psi(k) = a \exp^{-\left(\frac{k-b}{c}\right)^2}, \ \forall k \in [0, K] \tag{7}$$

As mentioned in BDM, it reduces the number of parameters substantially as ψ only requires 3 parameters instead of K for an unrestricted impulse response. However, this functional approximation approach involves one to take a stance on K.

In BDM, they use $K = 45$ in their codes. However, no explanation for this choice is given in the paper or elsewhere. This choice of 45 also seems arbitrary in light of the fact that the paper shows impulse response functions and multipliers for a horizon of 30 quarters, and pays particular attention to the multipliers at 20 quarters in the text.

We kept all details of their analysis the same and in their replication programs only changed the number of MA lags used to estimate the Gaussian function parameters that give rise to the government spending and output responses shown in Figure 3 of the paper. Our findings suggest that the main results, stating that the posterior probability of a multiplier to a contractionary shock is larger than an expansionary shock being above 0.95, are not robust to reducing the number of MA lags used for estimation. This is shown for the sample employed by BDM in the left panel of Figure 8. When we choose K=30, the posterior probability of a larger multiplier for contractionary shocks falls at all horizons, and notably never reaches 0.95. If we use $K = 20$, the posterior probabilities rise at shorter horizons. This suggests that posterior probabilities are not monotonically decreasing in the choice of K, but are sensitive to it.

Figure 8. FAIR results employing narrative identification scheme: choice of MA lags



Note: The figure shows the posterior probability that the cumulative multiplier for a contractionary shock is larger than an expansionary shock at a given horizon, h. The left panel shows results from the sample employed in Barnichon et al. (2022) code, which spans 1901q2-2015q4, and the right panel shows results for when we employ the full Ramey and Zubairy (2018) sample, which spans 1890q1-2015q4.

When we conduct the same analysis with the Ramey and Zubairy (2018) sample, as the paper claims to do, shown in the right panel of Figure 8, a similar picture emerges for K=30. There is weaker evidence of a larger multipliers for contractionary shocks. Notably, in this case when we consider $K = 20$, the posterior probability of a larger multiplier for contractionary shocks never exceeds 0.88, casting serious doubt on the baseline FAIR results of BDM for the narrative identification approach.

4 Ben Zeev's (2020) Quadratic Specification

In our quadratic specification we make use of the Jorda (2005) local projections method within a sign-dependent model, where a Bayesian estimation and inference procedure is performed by assuming a diffuse normal-inverse Wishart prior distribution for the local projection regressions' coefficients and residual variance. To account for correlations of the error term across time as well as outcome variables (i.e., detrended GDP and government spending), we apply a correction to the standard errors within our Bayesian estimation procedure, based on Driscoll and Kraay (1998) and following Auerbach and Gorodnichenko (2012)'s use of this correction in a classical setting, which accounts for arbitrary auto- and cross-correlations of the error term. In doing so we accord with the reasoning from Miranda-Agrippino and Ricco (2021), who estimate a hybrid VAR-local-projections model and follow the suggestion from Müller (2013) to increase estimation precision in the presence of a misspecified likelihood function (as in our and their setting) by replacing the original posterior's covariance matrix with an appropriately modified one.

We now turn to a general description of the estimation procedure.

4.1 Baseline Econometric Specification and Estimation

The estimation proceeds in two steps, which we conveniently discuss separately below.

Identification of Fiscal News Shock. The first step extracts the fiscal news shock by regressing the news series from Ramey and Zubairy (2018) on own four lags as well as four lags of the detrended GDP and government spending variables from Ramey and Zubairy (2018) .

Local Projection Regressions. The second step runs local projection regressions of detrended GDP and government spending variables on raw and squared values of the standardized residual from the first step. To reinforce the validity of our identification, we also add to these regressions four lags of the raw and squared values of the standardized residual as well as the detrended GDP and government spending variables. As to be explained below, we construct impulse responses to positive and negative fiscal news shocks from the first- and second-order polynomial coefficients of these regressions.

The econometric framework embodied in the two steps described above can be formally presented with the following the three-equation system:

$$
\begin{aligned}
(8) \qquad F_t &= \Gamma_{1,F}F_{t-1} + ... + \Gamma_{p,F}F_{t-p} + \Gamma_{1,GDP}GDP_{t-1} + \cdots \\
\cdots &+ \Gamma_{p,GDP}GDP_{t-p} + \Gamma_{1,G}G_{t-1} + ... + \Gamma_{p,G}G_{t-p} + C + \epsilon_t, \\
(9)\ GDP_{t+h} - GDP_{t-1} &= \alpha_{h,GDP} + \Xi_{h,GDP}\hat{\epsilon}_t + \Phi_{h,GDP}\hat{\epsilon}_t^2 + \Psi_{h,GDP}(L)Z_{t-1} + u_{t+h,GDP}, \\
(10) \qquad G_{t+h} - G_{t-1} &= \alpha_{h,G} + \Xi_{h,G}\hat{\epsilon}_t + \Phi_{h,G}\hat{\epsilon}_t^2 + \Psi_{h,G}(L)Z_{t-1} + u_{t+h,G},
\end{aligned}
$$

where F_t is the fiscal news series and GDP_t and G_t are the detrended GDP and government spending variables; $\Gamma_{i,j}$ (with $i = 1, ..., p$ and $j = F, GDP, G$) are scalar coefficients, where p denotes the number of lags which we set to 4 in accordance with the quarterly frequency of our data; C is a constant; ϵ_t is the true residual from Equation (8) (representing the true fiscal news shock in our setting); $\hat{\epsilon}_t$ is the estimated residual from Equation (8), thus representing our estimated fiscal news shock, and is normalized to have unit variance; the α's are constants; $\Psi_{h,GDP}(L)$ and $\Psi_{h,GDP}(G)$ are lag polynomial operators that introduce four lags of the variables comprising control vector Z ($\hat{\epsilon}_t$, $\hat{\epsilon}_t^2$, GDP_t, and G_t); and $\Xi_{h,GDP}$ ($\Xi_{h,G}$) and $\Phi_{h,GDP}$ ($\Phi_{h,G}$) are the first- and second-order effects of $\hat{\epsilon}_t$ on detrended GDP (detrended government spending), where $\Xi_{h,GDP} + \Phi_{h,GDP}$ ($\Xi_{h,G} + \Phi_{h,G}$) and $\Xi_{h,GDP} - \Phi_{h,GDP}$ ($\Xi_{h,G} - \Phi_{h,G}$) give the responses of detrended GDP (detrended government spending) at period $h+t$ to a positive and negative one standard deviation shock at time t, respectively. The fiscal multipliers at each horizon h for positive and negative shocks are defined as $M_{h,+} = \frac{\sum_{v=0}^{h} \Xi_{v,GDP}+\Phi_{v,GDP}}{\sum_{v=0}^{h} \Xi_{v,G}+\Phi_{v,G}}$ and $M_{h,-} = \frac{\sum_{v=0}^{h} \Xi_{v,GDP}-\Phi_{v,GDP}}{\sum_{v=0}^{h} \Xi_{v,G}-\Phi_{v,G}}$, respectively. Finally, we define asymmetry in impulse responses and fiscal multipliers as the differences between these objects for positive one standard deviation shocks and negative ones.

We estimate Equations (8), (9), and (10) jointly by applying the Bayesian estimation algorithm for strong block-recursive structure put forward by Zha (1999) in the context of block-recursive VARs, where the likelihood function is broken into the different recursive blocks. In our case, we have only two blocks, where the first consists of Equation (8) and the second corresponds to Equations (9) and (10). As shown in Zha (1999), this kind of block separation along with the standard assumption of a normal-inverse Wishart conjugate prior structure leads to a normal-inverse Wishart posterior distribution for the block-recursive Equation parameters. Following the suggestion from

Müller (2013) to increase estimation precision in the presence of a misspecified likelihood function (as in our setting owing to the auto- and cross-correlation embodied in $u_{t+h,GDP}$ and $u_{t+h,G}$), we apply a standard error correction based on Driscoll and Kraay (1998) which accounts for arbitrary auto- and cross- correlations of the latter error terms.[6]

Operationally, for each posterior draw of the coefficients from Equation (8), we collect the estimated residual from this equation ($\hat{\epsilon}_t$) and use its raw and squared values as the explanatory variables in Equations (9) and (10) (along with four lags of control vector Z) to form a posterior distribution of $\Xi_{h,GDP}$, $\Phi_{h,GDP}$, $\Xi_{h,G}$, and $\Phi_{h,G}$. We use the posterior draws of these coefficients to construct the sign-dependent responses of the outcome variable at horizon h to a positive and negative one standard deviation shock along with the corresponding fiscal multipliers (as explained above). We generate 2000 such posterior draws from which we are then able to estimate the median sign-dependent impulse responses and fiscal multipliers along with their posterior confidence bands.

4.2 Results

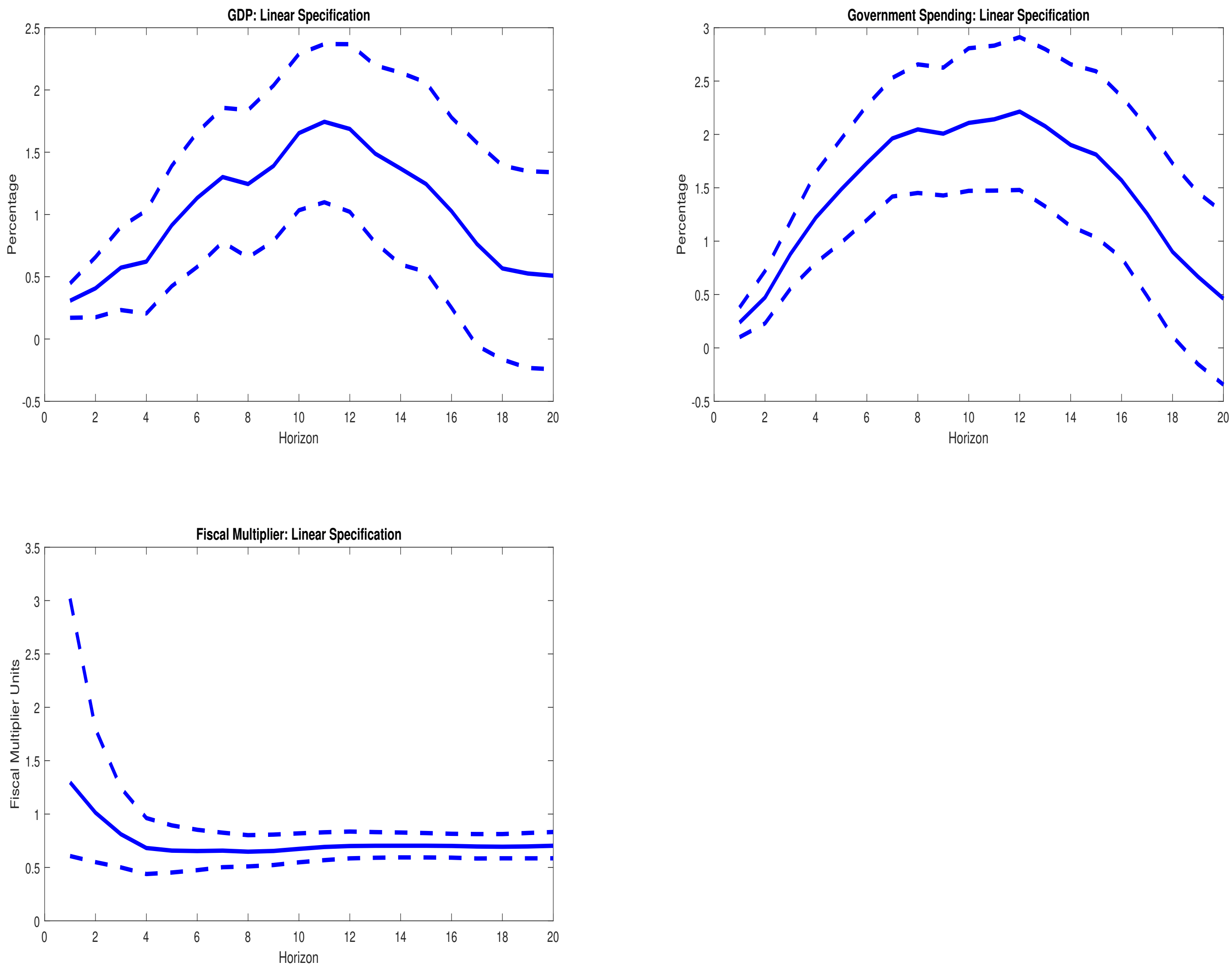
This section presents the main empirical results from the methodology described in the previous section. The first figure below shows the results from estimating Equations (9) and (10) without the terms relating to the squared value of the fiscal news shocks., i.e., it presents the linear responses of detrended GDP and government spending as well as their associated fiscal multiplier. The subsequent figures deal with sign-dependent responses and fiscal multipliers, with the first row of these figures depicting detrended GDP's response to a one standard deviation positive and negative fiscal news shocks as well as the difference between these responses; the second row showing identical objects for detrended government spending; and the third row presenting the fiscal multipliers with respect to positive and negative one standard deviation fiscal news shocks as well as the fiscal multiplier difference across these shocks. In all shown figures solid lines depict median estimates and dashed lines represent 95% posterior confidence bands.

6. To facilitate the ability of our inference procedure to account for the correlations between the residuals of detrended GDP and government spending, on top of these residuals' temporal correlations, we effectively treat these two outcome variables as cross-sectional units with heterogenous intercepts and slope coefficients.

Linear Impulse Responses. Prior to turning to the sign-dependent results, it is useful to consider also the linear impulse responses and fiscal multiplier obtained from estimating the linear analogues of Equations (9) and (10) (i.e., when omitting the terms relating to the squared value of the news shocks in these equations). Figure 9 presents the results from this exercise, which indicate a moderate fiscal multiplier that is significantly lower than one at business cycle frequencies. The multiplier reaches its peak of 1.1 on impact, thereafter leveling off at roughly two thirds.

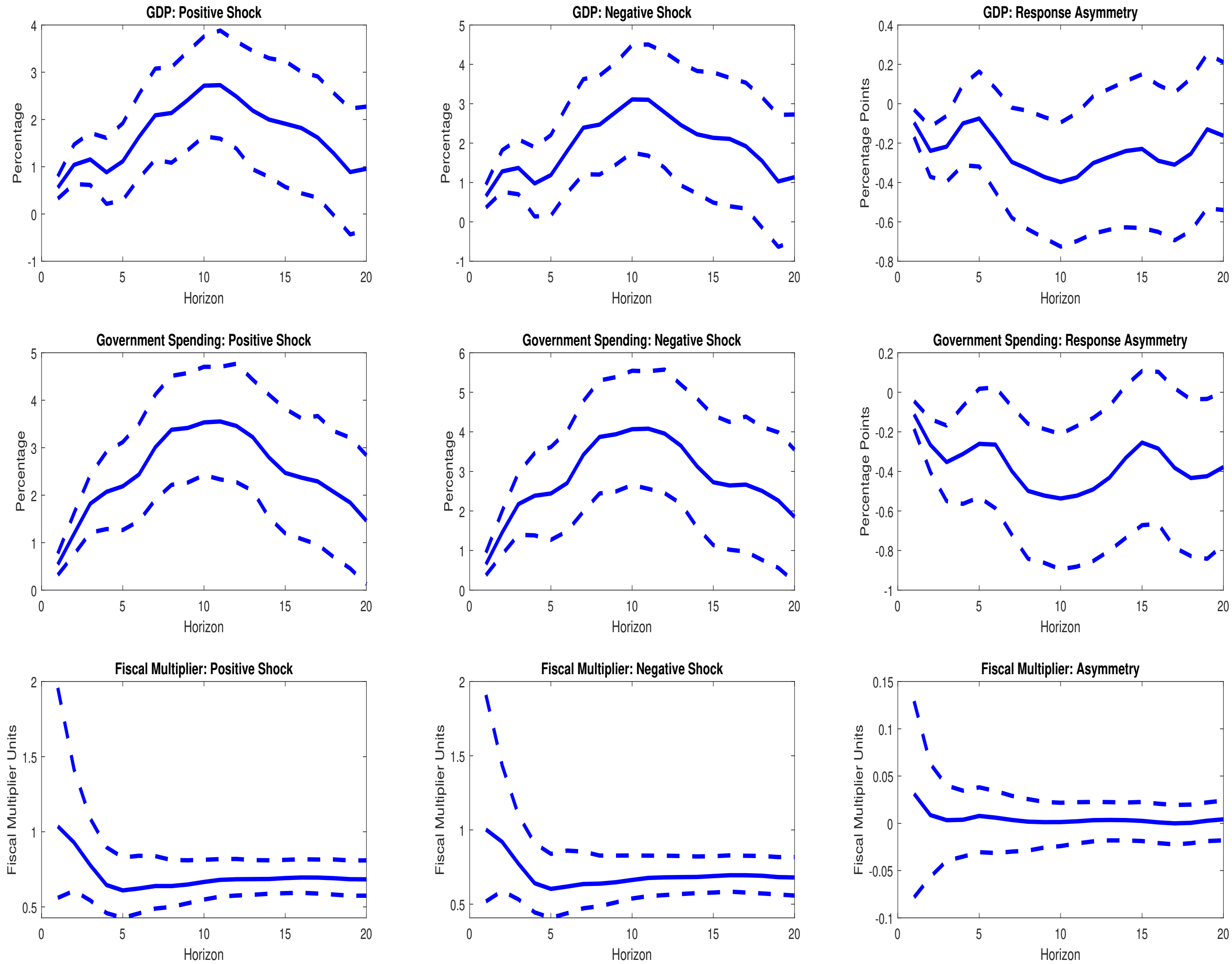
Sign-Dependent Impulse Responses. Figure 10 shows the results from estimation of System (8)-(10). The main takeaway from this figure is that there is no significant difference between the positive-shock- and negative-shock-induced fiscal multipliers for all considered horizons. The average estimated difference between the multipliers across horizons is 0.005, stressing that in terms of the point-wise estimates there is effectively no sign-dependency in the fiscal multiplier. In sum, the results from Figure 10 clearly fail to reject fiscal multiplier symmetry.

Figure 9. Linear Impulse Responses and Fiscal Multiplier



Notes: This figure presents the linear impulse responses of detrended GDP and government spending to a positive and negative one standard deviation fiscal news shocks as well as the linear fiscal multiplier associated with these responses. These results are obtained from estimation of the linear analogue of System (8)-(10), i.e., when excluding the terms related to the squared value of the news shock from Equations (9) and (10). Solid lines depict median estimates and dashed lines represent 95% posterior confidence bands. Horizon is in quarters.

Figure 10. Quadratic Specification: Sign-Dependent Impulse Responses and Fiscal Multiplier



Notes: This figure presents the sign-dependent impulse responses of detrended GDP and government spending to a positive and negative one standard deviation fiscal news shocks as well as the sign-dependent fiscal multiplier associated with these responses. These results are obtained from estimation of System (8)-(10). Responses to the negative shock are multiplied by -1. Solid lines depict median estimates and dashed lines represent 95% posterior confidence bands. Horizon is in quarters.

References

Auerbach, Alan J. and Yuriy Gorodnichenko, 2012. *Fiscal Multipliers in Recession and Expansion*, pp. 63–98. University of Chicago Press.

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.

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

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* .

Driscoll, John C and Aart C Kraay, 1998. "Consistent covariance matrix estimation with spatially dependent panel data." *Review of economics and statistics* 80(4): 549–560.

Gordon, Robert J. and Robert Krenn, 2010. "The End of the Great Depression: VAR Insight on the Roles of Monetary and Fiscal Policy." NBER Working paper 16380.

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

Jorda, Oscar, 2005. "Estimation and Inference of Impulse Responses by Local Projections." *American Economic Review* 95(1): 161–182.

Miranda-Agrippino, Silvia and Giovanni Ricco, 2021. "The Transmission of Monetary Policy Shocks." *American Economic Journal: Macroeconomics* 13(3): 74–107.

Montiel Olea, Jose Luis and Carolin Pflueger, 2013. "A Robust Test for Weak Instruments." *Journal of Business & Economic Statistics* 31(3): 358–369.

Müller, Ulrich K, 2013. "Risk of Bayesian inference in misspecified models, and the sandwich covariance matrix." *Econometrica* 81(5): 1805–1849.

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.

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.

Zha, Tao, 1999. "Block recursion and structural vector autoregressions." *Journal of Econometrics* 90(2): 291–316.