Investigating Government Spending Multiplier for the US Economy: Empirical Evidence Using a Triple Lasso Approach

Zacharias Bragoudakis[1,2,*] and Dimitrios Panas[3,4]

[1]*Bank of Greece.*

[2]*Department of Economics, National & Kapodistrian University of Athens.*

[3]*Tilburg School of Economics and Management.*

[4]*Systemic RM S.A., Athens, Greece.*

Abstract: An essential dilemma in economics that has yielded ambiguous answers is whether governments should spend more in recessions. This paper provides an extension of the work of Ramey & Zubairy (2018) for the US economy according to which the government spending multipliers are below unity, especially when the economy experiences severe slack. Nonetheless, their work suffered from some limitations with respect to invertibility and weak instrument problem.

The contribution of this paper is twofold: Firstly, it provides evidence that a triple lasso approach for the lag selection is a useful tool in removing the invertibility issues and the weak instrument problem. Secondly, the main results using a triple lasso approach suggest multipliers below unity for most cases with no evidence for differences between different states of the economy. Nevertheless, re-running the code in Ramey & Zubairy (2018), the case where WWII is excluded exhibits multipliers above unity, in both the military news and Blanchard-Perotti specifications, contradicting their baseline findings and providing evidence for a more effective government spending in recessions.

Keywords: Government spending, fiscal multipliers, debiased machine learning, triple lasso.
JEL Classification: C52, E62, H50, N42

1. INTRODUCTION

A main concern in fiscal policy analysis is the effect of government spending and more generally whether the effects of government spending on the economy depends on the state of the business cycle at the same time of the policy intervention. Is the government spending multiplier higher in recessions or not and what is its magnitude? Concerning the size of the multiplier, the literature is divided. In Keynesian models the government spending multipliers are nonlinear and state dependent. Using these models, some papers argue that government spending multiplier is different across states of the economy (Barro & Redlick (2011), Auerbach & Gorodnichenko (2012), Fazzari, Morley & Panovska (2015)) and in many cases above unity in recessions. Other researchers, using empirically driven approaches (Hauptmeier, Cimadomo & Kirchner (2010), Pereira & Lopez (2014), Ramey & Zubairy (2018)), argue that the multiplier is below unity and there is no evidence for difference across states. The different opinions basically stem from different assumptions implemented in model construction and different estimation techniques.

In this paper[1], the focus lies on the work of Ramey & Zubairy (2018). They estimate the government spending multipliers during periods of economic slack and when interest rates are close to the zero-lower bound. Moreover, they propose a not so novel method, estimating cumulative spending multipliers by a local projection method using instrumental variables. Their findings suggest that the government spending multipliers are below unity especially when the economy experiences high unemployment and they state that there is no evidence for differences in multipliers across states of slack. As they mention, their research faces some limitations. For example, due to narrative approach there is no controlled experiment. As Ramey & Zubairy (2018) mention, the military news variable and the Blanchard-Perotti shock that they use as instruments, are not informative about the size of the multiplier in cases where the government spending is about infrastructure. Moreover, they face a weak instruments problem.

A possible solution would be to select the lags in Ramey & Zubairy (2018) model in a parsimonious way, via triple lasso, which has never been used before in the literature. The goal is two-fold providing econometric and macroeconomic

*Address correspondence to this author at Econometric Forecasting Section, Economic Analysis And Research Department, 21, El.Venizelos Str.|GR 102 50|Athens, Greece; Tel: (+30) 210 320 3605; Fax. (+30) 210 320 3025; E-mail: ZBragoudakis@bankofgreece.gr

[1] An earlier version of this paper has been published in the Bank of Greece Working Paper Series, No. 292, October 2021.
DOI: https://doi.org/10.52903/wp2021292

contribution. The first is to investigate whether triple lasso is a useful tool in removing the invertibility issues and the weak instrument problem that Ramey & Zubairy (2018) face, while preserving consistency of the multipliers in every state. The second is to investigate whether proper shrinkage can reverse the results in Ramey & Zubairy (2018) and provide evidence for more effective government spending in recessions (higher multipliers).

The results obtained using a triple lasso approach indicate multipliers below unity and there is indication for less weak instruments in recessions but not in expansions. That suggests, the selection of the informative lags by triple lasso may not affect dramatically the magnitude of the multipliers. Moreover, there is no clear statistical evidence for difference in multipliers across states of slack. Only in Ramey & Zubairy (2018) when Blanchard-Perotti shock is used as instrument, then in most horizons the HAC and Anderson-Rubin p-values indicate differences in multipliers. It seems that the Blanchard-Perotti shock is stronger instrument than the military news variable, by looking at the effective F-statistics.

The structure of this paper is as follows. In section 2 there is a review of the work of Ramey & Zubairy (2018), presenting their main findings, describing their econometric approach and the problems that occur. In section 3 the econometric method we used is mentioned. Follows section 4 with the results. Moreover, a comparison is made between Ramey & Zubairy (2018) analysis and triple lasso analysis. Afterwards, several robustness checks are exhibited. The final section includes the main remarks, limitations and final thoughts.

2. REVIEW OF RAMEY & ZUBAIRY (2018)

2.1. Description

"Should government spending multipliers be higher in recessions?", is a question that Ramey & Zubairy (2018) try to answer for the US economy, gathering quarterly historical data from 1889-2015. They investigate whether this phenomenon is indeed the case during periods of economic slack or when interest rates are close to the zero-lower bound. Their paper contributes to the existing literature by pointing out the way multipliers should be calculated, providing also a comparison between normal times and bad ones.

They claim that the key contributing factor for high multipliers during economic downturns is the assumptions that are made in the data generating process. They state that as long as data-consistent assumptions are used, even in periods of slack, there is no evidence of high multipliers. At this point it should be mentioned that as a measurement of economic slack they use the unemployment rate threshold of 6.5 percent, following Owyang, Ramey & Zubairy (2013). Above that threshold the economy is in a slack state. Moreover, they use interest rates instead of unemployment rates as a robustness check. They also use two identification shocks in their analysis. The first is military news which is constructed that way in order to capture the political and military events and the second one is a Blanchard & Perotti (2002) shock, which is created straight out of the government spending series.

More specifically, Ramey & Zubairy (2018)[2] use Jordà (2005) local projection method for estimating the impulse responses and multipliers for their baseline. Furthermore, they introduce a different approach for computing the government spending multipliers. Their instrumental variable approach estimates the cumulative multipliers in a one-step instrumental variables regression. They consider the following model in the linear case:

$$\sum_{j=0}^{h} y_{t+j} = \gamma_h + \varphi_h(L)_{t-1} + m_h \sum_{j=0}^{h} \vartheta_{t+j} + u_{t+h} \text{, for h = 0,1, (2.1)}$$

where $\sum_{j=0}^{h} y_{t+j}$ is the cumulative real GDP divided by trend GDP, $\sum_{j=0}^{h} \vartheta_{t+j}$ is the cumulative real government spending divided by trend GDP, which is endogenous and instrumented by military news or Blanchard-Perotti shock, *w* is a vector of control variables, $\varphi_h(L)$ is a lag polynomial of order four. Moreover, for their baseline scenario they include in *w* lags of y_t, g_t and lags of military news. For the Blanchard-Perotti specification, the vector of control variables *w* includes only lags of y_t and lags of Blanchard-Perotti[3] shock. Moreover, they include a dummy variable to allow the estimation of a state dependent model, which indicate the state of the economy when is hit by a shock. Using $I_{t-1} \times shock_t$ and $(1-I_{t-1}) \times shock_t$ as instruments for the cumulative government spending, in expansion and in recession case respectively.

$$\sum_{j=0}^{h} y_{t+j} = I_{t-1}\left[\gamma_{A,h} + \varphi_{A,h}(L)\omega_{t-1} + m_{A,h}\sum_{j=0}^{h} g_{t+j}\right] + (1-I_{t-1})\left[\gamma_{B,h} + \varphi_{B,h}(L)w_{t-1} + m_{B,h}\sum_{j=0}^{h} g_{t+j}\right] + u_{t+h} \text{, for h = 0,1, (2.2)}$$

To continue, they discuss the importance of three major wars of WWI, WWII and the Korean war and the behavior of government spending multipliers during these periods. Concisely, in WWI the economy had an increasing GDP, there was a low unemployment rate and the interest rates were higher than the zero-lower bound. On the contrary, in WWII the US economy was in a more severe state of slack and the interest rates were at the zero-lower bound. During that time government spending increased by 35% of GDP. Last, in the Korean war the government spending was heavily increased. Interestingly, in WWI, private activity was partially crowded out by the government spending, while in the other two wars this is not observed in such a scale. Furthermore, Ramey & Zubairy (2018) point out that even if there is a great government spending (in magnitude) does not necessarily suggest a

[2] The data is from Ramey & Zubairy (2018) replication package which can be found at the following link, http://econweb.ucsd.edu/~vramey/research/Ramey_Zubairy_replication_codes.zip. For details on the data construction, the use of instruments, or the interpretation of α_h as government spending multipliers, one can visit the following site at https://econweb.ucsd.edu/~vramey/research.html, where there is a supplementary appendix, defense narrative, programs and data.

[3] The Blanchard-Perotti shock is equal with the current government spending.

higher government spending multiplier. In all these three war periods the magnitude of government spending rose approximately by 35%. To summarize, these war periods have in common high government spending but the multipliers are less than unity in all cases apart from their first quarters.

To conclude, they perform a variety of robustness checks for the slack estimates as well as for the zero lower bound estimates. For the first case, they change the threshold into a time varying one. They examine whether different samples are having different results. There are two samples with WWII exclusion and the other sample is post WWII. Moreover, they are adding controls for taxes. They also, increase the unemployment rate cutoff point. For the second case, they redefine the ZLB state period depending on a different basis of T-bill and they also include taxes and inflation as additional controls. In all these scenarios the results are the same and their baseline estimates are indicated to be robust. However, in only one setting they find multipliers higher than one. In this setting they do not include WWII in the sample and they use military news shock as an identification shock. Therefore, their results conclude to the fact that there is no evidence of multipliers higher than the unity when the US economy encounters high unemployment rates or when the US economy experiences zero lower bound interest rates in the full sample.

2.2. Identifying the Problem

Fig. (**4**) in Ramey & Zubairy (2018) indicates that there is a serious weak instrument problem for some horizons pertaining to recessions. This problem may explain why they cannot reject the null that government spending effectiveness is the same in recessions and expansions.

Many elements of w_t are close to collinear, which may drive the weak instrument results. To see why, note that by the Frisch-Waugh-Lovell theorem in (2.2): $\hat{m}_h = (S'M_wG)^{-1}(S'M_wY)^{-1}$, where $S = (shock_1,\dots,shock_T)'$, W is the matrix with rows $w_t^{'}Y$ is the vector with rows y_t, G is the vector with rows g_t and $M_w = I_T - W(W'W)^{-1}W'$, where I_T is the $T \times T$ identity matrix. Therefore, the properties of $\hat{m}_h$ are driven by the invertibility of *W'W*. When several lags are close to collinear, *W* is close to non-invertible, contaminating all the tests that use $\hat{m}_h$ including weak instruments tests. So, a possible solution is to select the lags in w_t in a parsimonious way, via triple lasso and check whether the weak instrument problem and the invertibility issues are removed, while preserving consistency of $\hat{m}_{A,h}$ *and* $\hat{m}_{B,h}$.

3. ECONOMETRIC APPROACH

3.1. Plain, Adaptive and OLS lasso

The idea of double or debiased machine learning (DML) is to estimate a parameter of interest, which may be a treatment or structural parameter, in the presence of an unknown function of many regressors, which is called nuisance parameter. That nuisance parameter is estimated using DML methods, such as the Least Absolute Shrinkage and Selection Operator, known as lasso. There is one parameter of interest such as a causal parameter or treatment effect parameter and the other parameters have to be controlled. The main focus lies on the structural parameters, on obtaining consistent estimators of the parameters of interest at the rate $\sqrt{n}$, where n is the sample size used to estimate the parameter of interest. Moreover, due to asymptotic normality of the structural parameter estimates, which in our case are impulse responses of output to government spending, tests can be conducted in order to check whether the treatment is significant or not. Last, if the treatment still depends on the controls this is where the omitted variable bias comes in, and that is exactly why double lasso is used for. The problem that double lasso fixes is the asymptotic endogeneity bias or omitted variable bias (Chernozhukov, Chetverikov, Demirer, Duflo, Hansen, Newey & Robins (2018)). The same thing applies for the implementation of triple lasso which is explained in the following section.

The plain Lasso estimator proposed by Tibshirani (1996) is defined as:

$$\hat{\beta}_L = \arg\underbrace{min}_{\beta}\left[||y - \sum_{j=1}^{p} x_j\,\beta_j||^2 + \lambda \sum_{j=1}^{p} |\beta_j| \right]$$

The quantity minimized above corresponds to the least squares objective function but penalized with $\lambda\sum_{j=1}^{p}|\beta_j|$. Lasso uses an l_1 penalty operator in the penalty term. The absolute value in the penalty term is capable of setting some coefficients to exactly zero and thus the method performs variable selection on top of estimation. For $\lambda = 0$, lasso does not shrink any coefficient while for increasing values of λ penalized coefficients shrink towards zero fitting a *sparse* model. Following Zou (2006), the plain lasso method might be inconsistent in variable selection and thus not offer any sizeable gain in model interpretability despite accurate predictions. To correct this specific drawback of the plain lasso, Zou (2006) introduced the adaptive lasso estimator as the following optimization problem:

$$\hat{\beta}_{AL} = \arg\underbrace{min}_{\beta}\left[||y - \sum_{j=1}^{p} x_j\,\beta_j||^2 + \lambda \sum_{j=1}^{p} w_j|\beta_j| \right]$$

In plain lasso, coefficients are penalized equally while the adaptive lasso introduces weights in the penalty term defined as

$w_j(\gamma) = \frac{1}{|\beta_j|^{\gamma}}$ for $\gamma > 0$ and $\hat{\beta}_j$ consistent pre-estimators of β_j.

This way, the adaptive lasso introduced by Zou (2006) corrects the inconsistencies of the plain method and tends to select the true non-zero coefficients in a sparse model.

Despite plain lasso being biased in finite samples, when the objective is just to predict (out-of-sample), it can be used reliably because some coefficients are important in the finite sample context. When focusing on in-sample interpretation and the interest is in the effects of particular variables and their values, adaptive lasso should be adopted.

Since plain lasso handles out-of-sample prediction relatively well, the tuning parameter should be chosen by Cross-

Validation. A grid of λ values is chosen and the CV error for each value of λ is computed. After that, the tuning parameter for which the CV error is minimized is selected. For the adaptive lasso that excels in in-sample, BIC is considered for the choice of λ. Even though the BIC can be computationally challenging, with adaptive lasso there is no need to check all the possible subsets and thus the BIC yields the tuning parameter without computational complications. However, using BIC is still controversial, due to the fact that it is not supported by any clear empirical results suggesting that in fact BIC yields the true model in the adaptive lasso context.

Nonetheless, adaptive lasso might be too parsimonious and for that reason there is the OLS post lasso approach. In OLS post lasso procedure, the model selection is accomplished by plain lasso and afterwards OLS is applied. That technique has as a result to make the estimates less biased (Belloni & Chernozhukov (2013)). It should be noted, that even though lasso performs both regularization and variable selection, by doing a post-lasso analysis the regularization step is "sacrificed" to ensure consistency of the multipliers.

3.2. The Baseline Model

A similar model as in Ramey & Zubairy (2018) is considered, for each horizon $h = 0,1\ldots,20$. For a single regime, the model is:

$$\sum_{j=0}^{h} y_{t+h} = \alpha_h \sum_{j=0}^{h} g_t + w_t^T \beta_h + u_{t+h}, \tag{3.1}$$

$$\sum_{j=0}^{h} g_t = \gamma shock_t + w_t^T \delta + v_t, \text{ for h = 0, 1, 2, …} \tag{3.2}$$

where y_t is real GDP divided by trend GDP, g_t is real government spending divided by trend GDP - which is endogenous and instrumented by military news or Blanchard-Perotti shock. The threshold variable is q_t, the first lag of the unemployment rate. The exogenous regressors w_t contain an intercept and four lags of y_t, g_t, $shock_t$.

Let $z_t = (shock_t, w_t)'$, $x_t = (g_t, w_t)'$, $\theta_h = (\alpha_h, \beta_h)'$, and impose: $E(z_t u_{t+h}) = 0$ (instrument validity), $E(z_t g_t')$ is full rank (instrument strength) and some weak dependence assumptions on the mean zero errors u_{t+h}, v_t (for example, α-mixing assumptions, as u_{t+h} are typically autocorrelated as indicated in Stock & Watson (2018)) (dependence assumptions).

Then the IV estimator $\hat{\theta}_h = \left(\sum_{t=1}^{T} z_t x_t'\right)^{-1}\left(\sum_{t=1}^{T} z_t y_t\right)$ is consistent for the true value θ_h, and $\hat{\alpha}_h$, the first element of $\hat{\theta}h$, is the estimated government spending multiplier at horizon h. Two regimes are considered; one with unemployment rate q_h below γ (expansion: $q_t \le \gamma$) and the second with unemployment rate above γ (recession, $q_t > \gamma$). Therefore, the model above is estimated for these two samples separately, and two estimates are obtained:

$$\hat{\theta}_h^L = \left(\sum_{t=1}^{T} z_t x_t' 1[q_t \le \gamma]\right)^{-1}\left(\sum_{t=1}^{T} z_t y_t 1[q_t \le \gamma]\right)$$

$$\hat{\theta}_h^H = \left(\sum_{t=1}^{T} z_t x_t' 1[q_t > \gamma]\right)^{-1}\left(\sum_{t=1}^{T} z_t y_t 1[q_t > \gamma]\right),$$

whose first elements are the government spending multipliers $\hat{\alpha}_h^L$ and $\hat{\alpha}_h^H$.

There is interest whether α_h is different for the two regimes $1[q_t \le \gamma]$ and $1[q_t > \gamma]$ and estimate the following two sets of moment conditions for each h:

IV: $E\left[z_t\left(y_{t+h} - x_t'\theta_h\right)\right] = 0$

$$E\left[z_t\left(y_{t+h} - x_t'\theta_h\right)1[q_t \le \gamma]\right] = 0$$

$$E\left[z_t\left(y_{t+h} - x_t'\theta_h\right)1[q_t > \gamma]\right] = 0$$

OLS: $E\left[x_t\left(y_{t+h} - x_t'\theta_h\right)\right] = 0$

$$E\left[x_t\left(y_{t+h} - x_t'\theta_h\right)1[q_t \le \gamma]\right] = 0$$

$$E\left[x_t\left(y_{t+h} - x_t'\theta_h\right)1[q_t > \gamma]\right] = 0$$

The first set of moment conditions is an IV regression for the linear, recession ($q_t < \gamma$) and expansion case ($q_t < \gamma$). The second set is an OLS regression, which is presented as Jordà's projection method (Jordà (2005)).

The goal is to run the aforementioned regressions for each h, with three thresholds of γ. Ramey & Zubairy (2018) use γ = 6.5 and γ = 8 as robustness check. The third value of γ would be equal to 8.3363, as estimated by Rothfelder & Boldea (2019). To continue, there are several specifications examined with different samples, controls and shocks. Furthermore, the first stage F-statistic (Kleibergen-Paap rk Wald F statistics) over h as well as the Anderson-Rubin p-values, will be computed and compared to the one when partialling out is not used. Also, the second stage effective F-statistic is shown. The difference of the first stage F-statistic and the valid 5 percent threshold based on Olea & Pflueger (2013) will be computed, plotted and compared to Ramey & Zubairy (2018) results. Last, two extra specifications are examined. In the first one the order of the polynomial is changed from four to five and eight. The second concerns changes in the bandwidth size. Both specifications are trying to mitigate the autocorrelation in the errors in Ramey & Zubairy (2018).

3.3. Triple Lasso

At this point, triple lasso procedure is described in details. At first, we run the code of Ramey & Zubairy (2018), obtaining the IV estimates of α_h using all the lags; that means four lags of y, g and the *shock* (military news or Blanchard-Perotti shock). Then, we subtract $\alpha_h \sum_{j=0}^{h} \vartheta_{t+j}$ from $\sum_{j=0}^{h} y_{t+j}$ and we regress this new variable performing the first lasso on the lags (w_t). Lasso will reveal which parameters are zero. Once we do that, the results for these β_h are stored. Next, we perform for the second time lasso for $\sum_{j=0}^{h} g_{t+j}$ on the lags (w_t).[4] After that, we regress the instrument which would be the military news variable or the Blanchard-Perotti shock on all these lags (w_t) performing for the third and final time lasso.

The important part is the orthogonalization as mentioned in Chernozhukov et al. (2018). The estimate of α_h must not depend on this selection, made by triple lasso. So, in order to ensure that the following procedure is performed: First, the part of $w_t^T \beta_h$, with the lasso estimates, is taken out of $\sum_{j=0}^{h} y_{t+j}$ and the residuals are obtained. Second, the part of $w_t^T \beta_h$, estimated by lasso, is subtracted from $\sum_{j=0}^{h} g_{t+j}$ and the residuals are also obtained. Last, the part of $w_t^T \delta$, with the lasso estimates, is taken out of the instrument of military news or Blanchard-Perotti shock and the residuals are obtained. Now there are three sets of residuals. So, the residuals from the first set are regressed on the second set with the third set of residuals becoming instrument.

It is worth noting that getting the first lasso of $\sum_{j=0}^{h} y_{t+j}$ on w_t, could just be a regression of $\sum_{j=0}^{h} y_{t+j}$ on g_t and w_t, not penalizing the α_h but only the lags (w_t). However, in this case, this procedure does not work. That is because in time series more lags are included in order to ensure instrument validity. Also, the valid moment condition is equal to zero, so that $E\left[z_t\left(y_{t+h} - x_t^{'}\theta_h\right)\right] = 0$ holds. So, if enough lags are included in Z_t it will hold. That does not necessarily mean to put all the lags. At this point, it is unknown which lags should be contained. Therefore, before the needed lags are actually selected, a consistent estimator of α_h is required. The consistency is necessary because if too few lags are put, then the instruments may be invalid and the α_h coefficient will be asymptotically biased, and thus inconsistent.

To avoid that the first step is performed as mentioned above. The α_h is obtained performing an IV estimate including all the lags. It is safe to assume that if all lags are included the estimator will not be the best but it will be consistent. Then $\alpha_h \sum_{j=0}^{h} g_{t+j}$ is subtracted from $\sum_{j=0}^{h} y_{t+j}$ and then continue performing the rest aforementioned steps.

4. RESULTS

4.1. Comparison with Ramey & Zubairy

The dilemma about higher multipliers in recessions has been addressed, among others, by Ramey & Zubairy (2018). Their results indicate no evidence for multipliers higher than unity. The results performing triple lasso are in line with their findings. In Table **4.1** the estimates of multipliers across states of slack are presented for both analyses. The left column describes the specification shock used as well as the two and four years integral. The other columns show the linear, high unemployment (recession) and low unemployment (expansion) cases. Inside the parentheses, the HAC robust standard errors are displayed. Observing Ramey & Zubairy (2018) results, the multipliers across states using the military news variable as instrument, show multipliers between 0.603 and 0.713 for both 2-years and 4-years horizon. Moreover, using the Blanchard-Perotti shock the multipliers for the 2-years horizon are 0.384, 0.680 and 0.304 for the linear, recession and expansion case respectively. For the 4-years integral, there is a marginal increase in multipliers comparing to the 2-years integral. Even though, the multipliers are below unity, there is evidence for differences between states of slack when looking at the HAC and Anderson-Rubin p-values indicating p-values that are significant at the 1% of confidence level. Both tests, examine whether there is difference between the multipliers or not. The only difference between HAC p-values and Anderson-Rubin p-values is that Anderson-Rubin is a weak instrument robust test, that means that even if there are weak instruments the p-values are correct as the sample size goes to infinity.

To continue, in triple lasso analysis the results are similar. In the second major column of Table **4.1** the estimates of multipliers across states of slack are illustrated, for the triple lasso specification. The multipliers in linear and recession case for the military news shock are similar to Ramey & Zubairy (2018) results but in expansion case the multipliers are almost half of what they find. Moreover, using the Blanchard-Perotti specification the multipliers for the linear case are similar in both analyses. However, in triple lasso analysis for the recession case the multipliers are extremely low but insignificant, almost one third of what Ramey & Zubairy (2018) find, 0.117 and 0.146 for the 2-years and 4-years horizon correspondingly, while in expansion case the multipliers are enormous and so their standard errors. Last, no difference across states is observed.

Furthermore, there are two similarities regarding Ramey & Zubairy (2018) and triple lasso analysis. In every specification, the initial multiplier in the linear case, is above one and then it starts falling. As explained in Ramey (2011), when there is information about a rise in future government spending, there is an instantaneous effect to GDP. Since GDP increases faster than government expenditures, the multiplier subsequently will be larger. The second similarity is observed when the analysis involves Blanchard-Perotti shock. The multipliers in that specification are lower compared to military news specification. There are two main reasons for that and both are elaborated in Ramey (2011). The first reason is that when shocks are expected, then the impulse responses will not capture the anticipatory rise in GDP. The second reason lies in the structure of Blanchard-Perotti shock. Blanchard-Perotti shock is constructed as the current government spending and consequently, there is correlation between the measurement error of the instrument and the measurement error of the government spending.

---

[4] This new variable is created because of endogeneity concerns. The cumulative g is endogenous. In order to correct the endogeneity problem, we subtract the cumulative g multiplied by the consistent estimator of α_h.

Table 4.1. Comparison.

	Ramey & Zubairy (2018)			Triple Lasso		
	Linear Model	**High Unemployment**	**Low Unemployment**	**Linear Model**	**High Unemployment**	**Low Unemployment**
Military news shock:						
2-years integral	0.664***	0.603***	0.595***	0.654***	0.527***	0.366***
	(0.067)	(0.095)	(0.091)	(0.227)	(0.114)	(0.134)
4-years integral	0.713***	0.682***	0.668***	0.745***	0.639***	0.347**
	(0.044)	(0.052)	(0.121)	(0.170)	(0.102)	(0.139)
Blanchard-Perotti shock:						
2-years integral	0.384***	0.680***	0.304***	0.390*	0.117	2.727
	(0.111)	(0.102)	(0.111)	(0.203)	(0.091)	(1.670)
4-years integral	0.474***	0.770***	0.348***	0.481**	0.146	2.677
	(0.110)	(0.075)	(0.107)	(0.207)	(0.102)	(2.843)

Note: The above table presents in the first major column Ramey & Zubairy (2018) results for the baseline scenario. The results for the following major column are obtained performing triple lasso analysis. There are three sub-columns presenting the linear case, the high unemployment or recession case and last the low unemployment or expansion case. The HAC-robust standard errors are inside the parentheses. The estimates for the 2-years integral and 4-years integral are the estimates for the seventh horizon and the fifteenth horizon respectively; that is because Ramey and Zubairy's data are quarterly constructed.

Concluding, the results for the baseline scenario are very close. The differences are probably because Ramey & Zubairy (2018) specify a polynomial of order four for every horizon. By implementing a lasso approach the structural model comprising both y_t, g_t and z_t it is a parsimonious VAR model, changing in every horizon. Only the important variables are kept in, capturing more accurately the variation.

4.2. Robustness Check

As in Ramey & Zubairy (2018), a lot of robustness checks are taking place in order to ensure the validity of their results. The same applies for this analysis. The estimates of multipliers across states of slack are tested when there is an additional control for taxes[5] when the sample changes by excluding the WWII[6] and when the unemployment cutoff point changes from 6.5 to 8 and 8.3633.

The results presented in Table **4.2** depict the robustness check comparison between Ramey & Zubairy (2018) and triple lasso analysis for the specifications which additional control for taxes is implemented and when WWII is excluded. In the upper panel where there is an additional control for taxes, the multipliers are in line with the baseline results of Table **4.1** presenting almost the same multiplier estimates. The results compared to Ramey & Zubairy (2018) are quite different. Only when the additional control for taxes is implemented the results are the same as the baseline case comparison. The low unemployment case using as instrument the Blanchard-Perotti shock, are unreliable due to very large multipliers and even greater standard errors. For both analyses there is no evidence of differences across states.

To continue, the lower panel in Table **4.2**, presents the multipliers when WWII is excluded[7]. The government spending multipliers present great changes and for some cases are above unity. In Table 3 of Ramey & Zubairy (2018) in high unemployment state they find multipliers of 0.72 and 0.89 for the 2-years and 4-years integral when using the military news variable. For the Blanchard-Perotti specification, they find 0.98 and 1.62. Running their code, the results are different obtaining for the military news variable multipliers of 1.774 and 1.345 and for the Blanchard-Perotti shock 1.017 and 1.644 for the 2-years and 4-years integral respectively. In Table **4.2** the results are depicted running their code and not what they present in their paper. So, there is evidence for multipliers above unity when the economy is in recession for both instruments, for every integral.

Comparing triple lasso results with Ramey & Zubairy (2018) and having military news as instrument, the multipliers are lower in every case apart from the linear where the results are similar. Looking at the high unemployment case, in both 2-years and 4-years integral the multipliers are 0.882 and 0.811 respectively. Nonetheless, the multipliers are more statistically significant and their standard errors are quite smaller than those of Ramey & Zubairy (2018). In contrast, when Blanchard-Perotti shock is used for the linear case, the multipliers obtained by triple lasso are almost four times larger and for the high unemployment state the multipliers are six and eight times smaller compared to what Ramey & Zubairy (2018) find.

[5] In Ramey & Zubairy (2018), they add lags of the average tax rate, given by the tax revenues as a ratio of GDP.

[6] Period of WWII that is excluded: 1941q3–1945q4.

[7] The sample with WWII exclusion is because United States increased excessively their military spending during WWII Gordon & Krenn (2010).

Additional robustness checks are examined when the unemployment cutoff point changes from 6.5 to 8 and 8.3633. In Table **4.3**, the results are presented for both analyses. The upper panel depicts the government spending multipliers when the unemployment cutoff point is 8 percent, while the lower panel when it is 8.3633. There are many changes when the cutoff point changes. In triple lasso approach when the military news shock is used the multipliers for the linear model are similar to the baseline. The recession case presents smaller multipliers 0.289 and 0.424 for the 2 and 4 years integral respectively, while the expansion case exhibits slightly larger multipliers compared to the baseline. Moreover, when the Blanchard-Perotti shock is used the multipliers in recession case are almost zero and in expansion almost 4.5. The results are not reliable because the multipliers are not statistically significant even at the 95% confidence level. Last, there are no significant differences in triple lasso results when the threshold changes from 8 to 8.3633.

Table 4.2. Robustness Check Comparison: including control for taxes and excluding WWII.

	Ramey & Zubairy (2018)			**Triple Lasso**		
	Linear Model	**High Unemployment**	**Low Unemployment**	**Linear Model**	**High Unemployment**	**Low Unemployment**
	Additional Control for Taxes					
Military news shock:						
2-years integral	0.671***	0.664***	0.563***	0.654***	0.545***	0.372***
	(0.068)	(0.136)	(0.088)	(0.226)	(0.127)	(0.121)
4-years integral	0.713***	0.694***	0.595***	0.745***	0.640***	0.360**
	(0.040)	(0.079)	(0.124)	(0.170)	(0.103)	(0.128)
Blanchard-Perotti shock:						
2-years integral	0.375***	0.674***	0.374***	0.390*	0.120	3.148
	(0.126)	(0.134)	(0.085)	(0.197)	(0.094)	(1.940)
4-years integral	0.445***	0.785***	0.395***	0.481**	0.150	2.988
	(0.151)	(0.138)	(0.116)	(0.191)	(0.106)	(3.215)
	Excluding WII					
Military news shock:						
2-years integral	0.772***	1.774*	0.557***	0.657**	0.882**	0.370***
	(0.201)	(0.955)	(0.154)	(0.271)	(0.356)	(0.126)
4-years integral	0.742***	1.345**	0.535**	0.743***	0.811***	0.371***
	(0.159)	(0.646)	(0.220)	(0.175)	(0.212)	(0.120)
Blanchard-Perotti shock:						
2-years integral	0.128	1.017***	0.127*	0.389***	0.164	1.211*
	(0.080)	(0.823)	(0.074)	(0.149)	(0.139)	(0.736)
4-years integral	0.151	1.644***	0.179*	0.477***	0.233	1.441
	(0.093)	(1.895)	(0.107)	(0.143)	(0.221)	(1.415)

Note: The above table presents in the first major column Ramey & Zubairy (2018) results for the case where additional control for taxes is implemented and for the case where WWII is excluded. The results for the following major column are obtained performing triple lasso analysis. There are three sub-columns presenting the linear case, the high unemployment or recession case and last the low unemployment or expansion case. The HAC-robust standard errors are inside the parentheses. The estimates for the 2-years integral and 4-years integral are the estimates for the seventh horizon and the fifteenth horizon respectively; that is because Ramey and Zubairy's data are quarterly constructed.

Table 4.3. Robustness Check Comparison: Different Unemployment Rate Thresholds.

	Ramey & Zubairy (2018)			**Triple Lasso**		
	Linear Model	**High Unemployment**	**Low Unemployment**	**Linear Model**	**High Unemployment**	**Low Unemployment**
	Unemployment rate threshold of 8%					
Military news shock:						
2-years integral	0.664***	0.797***	0.563***	0.654***	0.289***	0.472**

	(0.067)	(0.205)	(0.088)	(0.227)	(0.068)	(0.202)
4-years integral	0.713***	0.760***	0.595***	0.745***	0.424***	0.438**
	(0.044)	(0.099)	(0.124)	(0.170)	(0.072)	(0.174)
Blanchard-Perotti shock:						
2-years integral	0.384***	0.639***	0.374***	0.390*	-0.003	4.212
	(0.111)	(0.095)	(0.085)	(0.203)	(0.103)	(3.231)
4-years integral	0.474***	0.690***	0.395***	0.481**	0.044	4.651
	(0.110)	(0.085)	(0.116)	(0.207)	(0.129)	(6.658)
	Unemployment rate threshold of 8.3633%					
Military news shock:						
2-years integral	0.664***	0.659***	0.602***	0.654***	0.287***	0.482**
	(0.067)	(0.201)	(0.092)	(0.227)	(0.064)	(0.205)
4-years integral	0.713***	0.712***	0.638***	0.745***	0.427***	0.447**
	(0.044)	(0.103)	(0.106)	(0.170)	(0.073)	(0.171)
Blanchard-Perotti shock:						
2-years integral	0.384***	0.352***	0.615***	0.390*	0.005	4.033
	(0.111)	(0.119)	(0.120)	(0.203)	(0.111)	(3.024)
4-years integral	0.474***	0.426***	0.663***	0.481**	0.053	4.712
	(0.110)	(0.118)	(0.090)	(0.207)	(0.143)	(6.930)

Note: The above table presents in the first major column Ramey & Zubairy (2018) results for the case where the unemployment rate threshold changes to 8% and 8.3633%. The results for the following major column are obtained performing triple lasso analysis. There are three sub-columns presenting the linear case, the high unemployment or recession case and last the low unemployment or expansion case. The HAC-robust standard errors are inside the parentheses. The estimates for the 2-years integral and 4-years integral are the estimates for the seventh horizon and the fifteenth horizon respectively; that is because Ramey and Zubairy's data are quarterly constructed.

Concluding, when the unemployment cutoff point changes from 8.0 and 8.3633 percent for the Ramey and Zubairy (2018) analysis, the results are pretty much the same. The only difference that can be observed is in the Blanchard-Perotti specification, in the recession case the multipliers are lower and in expansion larger when the cutoff point is 8.3633 compared to the 8 percent case.

4.3. Instruments Relevance

In Ramey & Zubairy (2018), the relevance of military news variable is under question. By construction this variable captures the changes of military spending by the government and thus should be exogenous to the economy. For that reason, they examine the relevance of the instrument using F-statistics and the Olea & Pflueger (2013) effective F-statistics and thresholds. Their results are illustrated in Fig. (**4**) of their analysis.

Briefly, in Fig. (**4**) of Ramey & Zubairy (2018), three cases are examined; full sample, post WWII, excluding WWII. It is clear, that military news variable at least for the initial horizons has relevance issues, while the Blanchard-Perotti shock remains strong. For the plots in the high unemployment case, it is observed that near the tenth horizon the Blanchard-Perotti shock drops below the threshold. Nonetheless, the Blanchard-Perotti shock may be an invalid instrument, as explained in Ramey & Zubairy (2018). Consequently, both instruments face relevance problems.

In Fig. (**4.1**), a similar approach is used to illustrate the test of instrument relevance across states of slack, using triple lasso. The baseline model is shown, as well as, the cases where additional control for taxes is introduced and where WWII is excluded. In linear and in expansion case the military news instrument is weak, remaining persistently below the threshold. The Blanchard-Perotti shock presents weak instrument issues for the low unemployment case and for the linear case, it shows high relevance up to the seventh horizon and then gradually drops, reaching below the threshold at the last horizons. In contrast, both instruments are strong for the recession case. The military news variable may start below the threshold, nonetheless, within a few horizons it rises rapidly.

It is worth noting that there are strong concerns that the critical value of 21.1 might not be valid for triple lasso. In Chernozhukov et al. (2018) is argued that double lasso selection should not affect, under reasonable assumptions, the properties of the IV IRF estimator, and that it is efficient. However, the selection procedure may affect the weak instrument critical value of Olea & Pflueger (2013), and it would be of interest to derive the asymptotic properties of the test. Nevertheless, no matter what the critical value is, it is probably large enough so that values of the effective F-statistic around 3, as observed in the linear and low unemployment cases, are still indications of weak instruments.

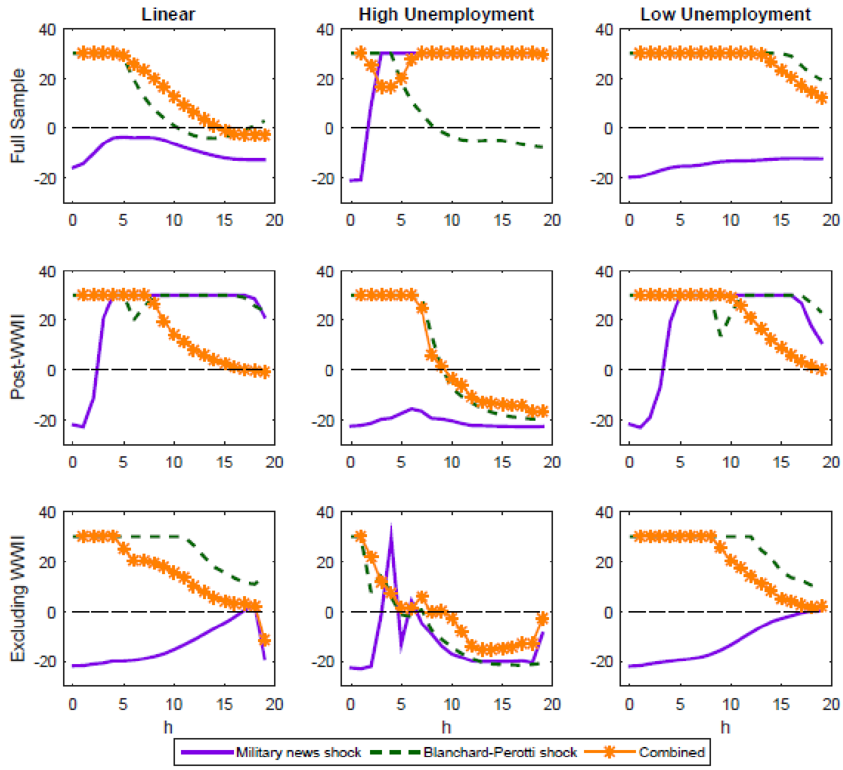

Fig. (4). Tests of instrument relevance across states of slack from Ramey & Zubairy (2018).

Note: "Slack" is when the unemployment rate exceeds 6.5 percent. The lines show the difference between the effective F-statistic and the relevant threshold for the five percent level, and are capped at 30. The effective Fstatistics are from the regression of the sum of government spending through horizon h on the shock at t and all the other controls from the second stage, separately for the military news variable (solid line), the Blanchard-Perotti shock (dashed line) and both instruments (line with asterisks). The first column shows the linear case, the second column shows the high unemployment state and the last column shows the low unemployment state. The full sample is 1890:1-2015:4, and the post-WWII sample spans 1947:3 - 2015:4.

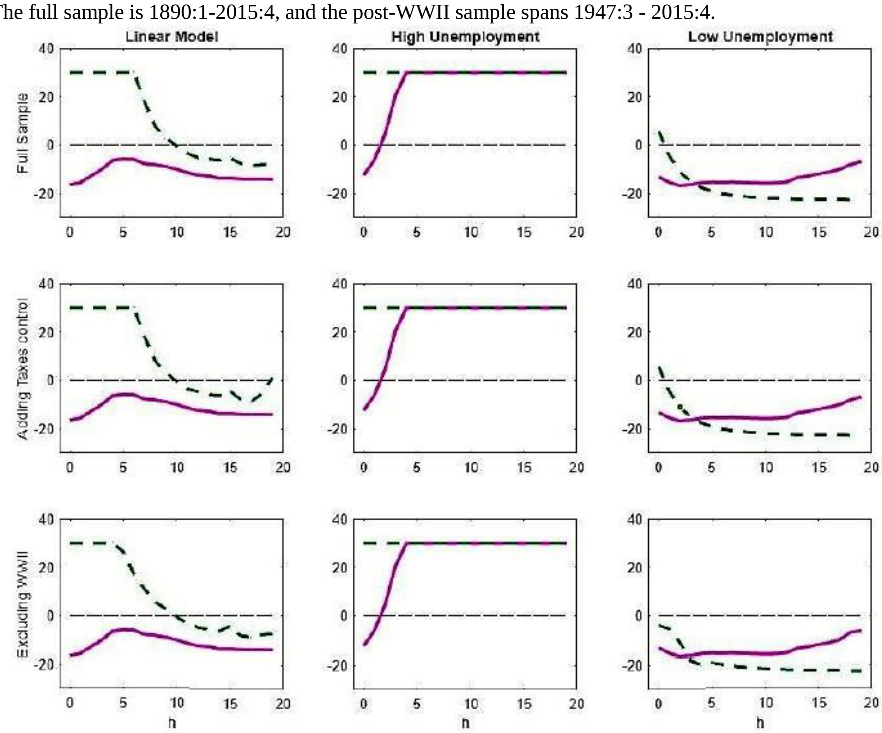

Fig. (4.1). Tests of instrument relevance across states of slack for triple lasso.

Note: The above graphs depict the instrument relevance across states of slack using triple lasso. That is the difference of the effective F-statistic and the valid 5 percent threshold based on Olea & Pflueger (2013). The threshold is 23.1 percent for the 5 percent critical value. for the linear case, the high unemployment case (recession) and the low unemployment case (expansion). The Blanchard-Perotti shock is presented by the dashed line and the military news variable is presented by the solid line. There are three specifications. The full sample is 1890q1–2015q4, the full sample adding controls for taxes and excluding WWII.

4.4. Additional Specifications

In the analysis of Ramey & Zubairy (2018), the Olea & Pflueger (2013), effective F-statistics and thresholds are used, because of the serially correlated errors in Jordà's method. They observe serial correlation even at horizon 0 and therefore they use automatic bandwidth according to Newey & West (1994). After running their code, the bandwidth is examined for every horizon. It is observed that the automatic bandwidth is very large, even greater than 28 for some cases. So, the additional specifications check that are considered for Ramey & Zubairy (2018) analysis, is first to examine the case were there is no automatic bandwidth but a fixed one in order to correct the standard errors of the multipliers. The second is to include an additional lag to their analysis in order to solve the autocorrelation problem as mentioned in Montiel Olea & Plagborg-Møller (2020). Last, an additional specification is tested. Instead of introducing a fifth lag, eight lags are introduced, aiming for a solution to the autocorrelation in the errors.

For the first case, an arbitrary a fixed bandwidth of 4 and later of 7 is used. The reason for the particular choices of 4 and 7 are based on the PACF (partial autocorrelation function) and ACF (autocorrelation function). The PACF indicates the correlation of the residuals with spikes, at which lag there is additional autocorrelation. The ACF indicates the autocorrelation of the series with their lagged values, it plots the decay. So, the goal is to include enough lags to capture that decay. When the decay goes below significant then there is no need to include more lags because more noise will be included.

So, the analysis is performed exactly as in Ramey & Zubairy (2018), but with different bandwidths. This is examined for the baseline scenario, the case where additional control for taxes is implemented and last for the case where WWII is excluded. The only features that change is the robust standard errors, the HAC p-values, the Anderson-Rubin p-values and the F-statistics for the instruments.

As seen from Table **4.4** using the military news as instrument, as the size of bandwidth shrinks from automatic (which is above 25 for every horizon), to 7 and to 4 the HAC and Anderson-Rubin p-values get larger. Nonetheless, this increase is marginal. Moreover, in Table **4.5** the military news variable gets weaker for every state of the economy, as the effective F-statistic drops when a fixed bandwidth is implemented. When the Blanchard-Perotti shock is used as an instrument, and the analysis is run with a 7 and 4 bandwidth imposition, the same marginal increase is noted for the HAC and Anderson-Rubin p-values. Moreover, as observed in the effective F-statistic, the results are ambiguous. For the first ten horizons the linear and the recession case, show a rise when the bandwidth is smaller, indicating that the instrument gets stronger. For the last ten horizons the effective F-statistic gets smaller as the bandwidth gets smaller. The same patterns are also noticed when an additional control for taxes is implemented as well as in the case where WWII is excluded.

Table 4.4. HAC and AR p-Values with Different Bandwidths.

	Military news as instrument						**Blanchard-Perotti shock as instrument**					
	HAC			**AR**			**HAC**			**AR**		
h	bw(auto)	bw(4)	bw(7)	bw(auto)	bw(4)	bw(7)	bw(auto)	bw(4)	bw(7)	bw(auto)	bw(4)	bw(7)
0	0.044	0.201	0.163	0.225	0.194	0.171						
1	0.042	0.058	0.049	0.238	0.243	0.238	0.517	0.595	0.573	0.536	0.597	0.579
2	0.000	0.006	0.002	0.242	0.243	0.246	0.084	0.158	0.129	0.188	0.188	0.174
3	0.008	0.070	0.046	0.253	0.303	0.294	0.031	0.075	0.053	0.123	0.108	0.097
4	0.261	0.435	0.394	0.387	0.533	0.508	0.017	0.054	0.034	0.103	0.085	0.076
5	0.523	0.651	0.621	0.573	0.688	0.667	0.010	0.041	0.024	0.088	0.071	0.063
6	0.811	0.859	0.846	0.818	0.864	0.852	0.008	0.043	0.023	0.078	0.072	0.061
7	0.954	0.964	0.961	0.954	0.964	0.961	0.005	0.048	0.025	0.070	0.078	0.063
8	0.823	0.857	0.847	0.819	0.856	0.843	0.004	0.045	0.022	0.061	0.072	0.058
9	0.967	0.973	0.971	0.966	0.973	0.971	0.004	0.049	0.024	0.055	0.076	0.060
10	0.870	0.890	0.882	0.872	0.890	0.884	0.003	0.049	0.024	0.048	0.074	0.058
11	0.791	0.815	0.806	0.798	0.818	0.811	0.002	0.045	0.021	0.040	0.070	0.053
12	0.814	0.830	0.823	0.820	0.831	0.828	0.001	0.034	0.014	0.032	0.056	0.041
13	0.894	0.899	0.896	0.895	0.899	0.898	0.001	0.029	0.012	0.028	0.049	0.036
14	0.991	0.991	0.991	0.991	0.991	0.991	0.001	0.028	0.011	0.030	0.048	0.036

15	0.924	0.926	0.926	0.924	0.926	0.925	0.001	0.026	0.010	0.032	0.048	0.037
16	0.898	0.901	0.901	0.897	0.901	0.900	0.000	0.024	0.010	0.036	0.048	0.040
17	0.833	0.834	0.835	0.830	0.836	0.834	0.000	0.024	0.010	0.041	0.052	0.045
18	0.766	0.763	0.766	0.763	0.768	0.766	0.001	0.027	0.012	0.048	0.060	0.052
19	0.755	0.750	0.753	0.752	0.757	0.754	0.000	0.026	0.011	0.049	0.064	0.055
20	0.771	0.765	0.767	0.770	0.772	0.768	0.000	0.025	0.011	0.047	0.067	0.057

Note: The HAC and Anderson-Rubin p-values with different bandwidths are examined for the baseline case in Ramey & Zubairy (2018).

Table 4.5. The Effective F-Statistic with Different Bandwidths.

	Military News as Instrument									**Blanchard-Perotti Shock as Instrument**								
	FKplin			**FKprec**			**FKpexp**			**FKplin**			**FKprec**			**FKpexp**		
h	bw(auto)	bw(4)	bw(7)	bw(auto)	bw(4)	bw(7)	bw(auto)	bw(4)	bw(7)	bw(auto)	bw(4)	bw(7)	bw(auto)	bw(4)	bw(7)	bw(auto)	bw(4)	bw(7)
0	7.33	3.46	3.99	2.03	2.89	2.71	3.35	1.98	2.31									
1	8.94	4.46	5.04	2.30	2.59	2.69	3.66	2.49	2.75	307.42	457.20	447.95	161.64	293.84	250.36	513.16	466.37	499.66
2	12.86	6.62	7.44	32.29	21.80	25.09	4.74	3.37	3.58	116.01	176.43	172.30	88.18	171.00	138.18	155.94	153.96	158.59
3	16.93	8.75	10.21	103.42	54.69	64.11	6.14	4.09	4.51	87.47	133.55	128.94	68.39	133.61	110.13	137.09	119.52	128.20
4	19.22	10.17	12.41	186.19	70.07	85.84	7.28	4.54	5.37	67.78	107.08	101.45	53.24	90.33	79.79	123.54	98.66	113.18
5	19.57	10.91	13.51	271.31	87.02	108.36	7.87	4.68	5.88	51.68	80.20	75.15	41.56	59.29	55.73	99.96	69.60	85.92
6	19.26	11.40	14.10	343.92	117.28	146.75	7.94	4.69	6.10	42.56	64.21	59.45	33.96	41.78	40.99	89.53	53.57	71.34
7	19.39	11.96	14.72	403.28	139.39	174.13	8.37	4.94	6.52	35.98	52.94	47.96	29.04	31.91	32.14	82.81	43.42	60.34
8	19.24	12.39	15.03	444.01	166.67	204.24	9.06	5.21	6.94	30.61	43.83	38.66	24.36	25.30	25.64	75.12	36.20	50.34
9	18.35	12.69	15.02	377.93	133.34	160.45	9.67	5.49	7.28	27.09	37.76	32.46	21.66	21.56	21.93	71.22	31.56	43.33
10	16.86	12.74	14.62	284.31	88.53	106.61	9.93	5.70	7.43	24.19	33.11	27.83	19.68	19.10	19.47	68.60	28.13	37.88
11	15.44	12.71	14.10	212.39	61.59	74.58	10.00	5.92	7.54	21.77	29.44	24.25	18.42	17.39	17.74	65.16	25.71	33.97
12	14.23	12.55	13.53	175.37	47.18	57.61	10.09	6.12	7.65	20.13	26.74	21.74	17.95	16.00	16.40	63.96	23.87	31.28
13	13.11	12.22	12.86	153.62	40.02	49.42	10.33	6.32	7.83	19.28	25.05	20.23	18.09	14.31	14.94	61.31	22.00	28.78
14	12.06	11.77	12.11	144.31	35.87	44.81	10.62	6.50	8.02	19.11	24.05	19.39	18.21	12.27	13.12	58.61	20.04	26.29
15	11.22	11.40	11.51	130.20	31.77	40.09	10.85	6.60	8.17	19.70	23.89	19.27	18.12	10.36	11.37	55.12	18.56	24.44
16	10.73	11.19	11.13	114.88	27.07	34.39	10.97	6.62	8.21	20.88	24.18	19.62	17.44	8.88	9.94	52.01	17.29	22.89
17	10.51	11.13	10.94	96.38	22.33	28.54	10.96	6.58	8.13	22.34	24.63	20.19	16.66	7.77	8.84	48.25	16.09	21.27
18	10.47	11.19	10.89	82.78	18.51	23.79	10.91	6.51	8.03	24.04	25.17	20.91	15.93	7.00	8.06	45.04	15.12	20.00
19	10.47	11.27	10.88	70.40	15.58	20.16	10.82	6.38	7.85	26.05	25.81	21.79	15.48	6.41	7.51	42.40	14.49	19.09
20	10.43	11.30	10.80	61.91	13.44	17.49	10.67	6.20	7.61	28.28	26.15	22.59	15.10	5.87	7.05	39.90	13.79	18.13

Note: The above results depict the effective F statistics with different bandwidths using triple lasso for the baseline case. "Fklin" column shows the effective F statistics for the linear case, "Fkrec" for the recession case (high unemployment) and "Fkexp" for the expansion case (low unemployment).

To continue, it is worth noticing that the HAC p-values and the Anderson-Rubin p-values when Blanchard-Perotti shock is used, present for almost every horizon significant p-values for the 1% confident level for both baseline scenario and also when additional control for taxes is implemented. That means for almost every horizon there is difference across states of slack. The government spending multipliers are below unity, but still in high unemployment state are different compared to the low unemployment state.

Furthermore, in Ramey & Zubairy (2018) analysis, they specify a polynomial of order four. If the structural model comprising both y_t, g_t and z_t would indeed be a VAR(4), then Montiel Olea & Plagborg-Møller (2020), claim that an extra lag would solve the autocorrelation problem. So, the other additional specification check that is performed here involves an additional fifth lag. That means, the z_t vector of control variables that is described at section 2, now includes 5 lags of y_t, g_t and 5 lags of the instrument used. When a fifth lag is added, the estimates of the multipliers also change providing slightly larger multipliers and relatively smaller standard errors compared to the baseline model of Ramey & Zubairy (2018). Still, when the correlograms are plotted, there are spikes indicating the presence of serial autocorrelation. Moreover, the HAC and Anderson-Rubin p-values state that the null hypothesis of no differences in multipliers across states of slack can be safely rejected. There is clear evidence that the multipliers are different in recession compared to those in expansion case. Also, the same can be said for the first 4 horizons when military news is used as an instrument.

Last, Ramey & Zubairy (2018) analysis is performed with 8 lags for the baseline, followed by an analysis with triple lasso. Once again, is examined whether adding more lags removes the autocorrelation problem. Still, the autocorrelation in the errors is there, which means that VAR(p) representation does not hold or the model is misspecified.

Overall, the addition of a fifth lag, does not solve the autocorrelation problem in the errors, as claimed Montiel Olea & Plagborg-Møller (2020), unless a VAR(4) is not the correct specification for the structural model involving y_t, g_t and the instrument z_t. The spikes at the correlograms are quite persistent and by including the fifth lag, some spikes may perish, but at the initial horizons are still present. This indicates that the initial model does not originate from a VAR(4) and a solution is yet to be found for correcting the autocorrelation. Nonetheless, these results do indicate that, at least for longer horizons of 4-5 years, the claims in Ramey & Zubairy (2018) regarding the lack of difference in multipliers across states of slack, and about multipliers not being above unity in recessions, are sensitive to the choice of lags and the accuracy of the HAC correction. Concluding, in all cases as the bandwidth gets smaller, the situation gets worse and the standard errors do not get any better. When the additional fifth lag and the eight lags specification is examined, the problem of autocorrelation is still present.

5. CONCLUSION

Ramey & Zubairy (2018) investigate whether government spending multipliers are higher during recessions. Their findings suggest that the government spending multipliers are below unity especially when the economy experiences severe slack and that there is no evidence for differences in multipliers across states of slack. Nonetheless, they encounter invertibility and weak instrument problems. This paper, investigates whether the lags in Ramey & Zubairy (2018) model can be selected in a parsimonious way. Robustness checks are examined in cases where additional control for taxes is implemented, when WWII is excluded and when the unemployment threshold changes from 6.5 percent to 8 and 8.3633. Additional specifications are also tested in order to correct the standard errors of the multipliers as well as the autocorrelation issues in Ramey & Zubairy (2018).

The use of triple lasso suggests that government spending multipliers are below unity. Government expenditure may provide military preparedness and deterrence against possible aggression and some could argue that this establishes economic stability but perhaps of the nature of military spending, being non-tradeable, it does not affect the GDP much. Moreover, using triple lasso shows that the problem of invertibility can be solved, while preserving consistency of the multipliers in every state. The weak instrument problem is not entirely solved, but there is indication that the instrument relevance is higher in high unemployment case. Moreover, the Blanchard-Perotti shock seems to be a stronger instrument compared to the military news variable, as shown by the effective F-statistics.

Furthermore, it is interesting to note that the sample with the WWII exclusion indicates higher multipliers. Particularly for the recession case with unemployment rate above 6.5 percent, multipliers are above unity at least in Ramey & Zubairy (2018) in both military news and Blanchard-Perotti specification, contradicting their baseline findings. That means more government spending does not necessarily mean larger multipliers. In periods of severe recession like WWII, nothing can help. Even if tremendous amounts are spent by the government, it is not spent on consumption or investments but in production of military equipment and generally military expenses. So, perhaps WWII is a special case and the multipliers for that period are extremely small and when the entire sample is examined then the average of the two period samples makes the multipliers smaller. Unfortunately, that cannot be checked due to the lack of observations.

As mentioned in Ramey & Zubairy (2018) both of their instruments cannot explain what the magnitude of the government spending multipliers is, in cases where the focus is not about military expenses but infrastructure. It should be noted however that is quite difficult to construct good instruments not related to military expenses. Moreover, even if the Blanchard-Perotti shock is a stronger instrument compared to military news using triple lasso, that does not imply that the weak instrument problem that Ramey & Zubairy (2018) face is vanished. Furthermore, using triple lasso may not be the only possible solution and further research is required. The rationale of this approach is model selection and it solves the invertibility issues. The lags are correlated so it is unclear what other methods would work. For example, Ridge regression could be used, but it would introduce bias that could persist asymptotically. Lasso introduces also bias but when the objective is just to predict (out - of sample), it can be used reliably because some coefficients are important in the finite sample context. Also, other lasso alternatives could be used, such as adaptive lasso or post OLS lasso, but might be too parsimonious leading to even smaller multipliers. Concluding, further research is required in order to investigate whether there are other proper shrinkage methods that can provide evidence for a more effective government spending in recessions.

ACKNOWLEDGEMENTS

The authors would like to thank S. Anyfantaki, H. Balfoussia, J. Joyce and G. Tavlas for helpful comments and suggestions. The paper has also benefited from the comments of the participants of the 25th International Conference on Macroeconomic Analysis and International Finance (ICMAIF) organized by the Department of Economics of the University of Crete in Rethymno, 29-31 July 2021. The paper has also been published in the bank of Greece as a working paper.

REFERENCES

Auerbach, A. J., & Gorodnichenko, Y. (2012). Measuring the output responses to fiscal policy. *American Economic Journal: Economic Policy*, *4*(2), 1-27.

Barro, R. J., & Redlick, C. J. (2011). Macroeconomic effects from government purchases and taxes. *The Quarterly Journal of Economics*, *126*(1), 51-102.

Belloni, A., & Chernozhukov, V. (2013). Least squares after model selection in high-dimensional sparse models. *Bernoulli*, *19*(2), 521-547.

Blanchard, O., & Perotti, R. (2002). An empirical characterization of the dynamic effects of changes in government spending and taxes on output. *the Quarterly Journal of economics*, *117*(4), 1329-1368.

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., & Robins, J. (2018). Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal*, *21(1), C1-C68.*

Fazzari, S. M., Morley, J., & Panovska, I. (2015). State-dependent effects of fiscal policy. *Studies in Nonlinear Dynamics & Econometrics*, *19*(3), 285-315.

Gordon, R. J., & Krenn, R. (2010). The end of the great depression 1939-41: Policy contributions and fiscal *multipliers (No. w16380). National Bureau of Economic Research.*

Hauptmeier, S., Cimadomo, J., & Kirchner, M. (2010). Transmission of government spending shocks in the euro area: Time variation and driving forces. *(No. 1219). European Central Bank.*

Jordà, Ò. (2005). Estimation and inference of impulse responses by local projections. *American economic review*, *95*(1), 161-182.

Knight, K., & Fu, W. (2000). Asymptotics for lasso-type estimators. *Annals of statistics*, 1356-1378.

Newey, W. K., & West, K. D. (1994). Automatic lag selection in covariance matrix estimation. *The Review of Economic Studies*, *61*(4), 631-653.

Olea, J. L. M., & Pflueger, C. (2013). A robust test for weak instruments. *Journal of Business & Economic Statistics*, *31*(3), 358-369.

Olea, J. L. M., & Plagborg-Møller, M. (2020). Local Projection Inference is Simpler and More Robust Than You Think. *arXiv preprint arXiv:2007.13888*.

Owyang, M. T., Ramey, V. A., & Zubairy, S. (2013). Are government spending multipliers greater during periods of slack? Evidence from twentieth-century historical data. *American Economic Review*, *103*(3), 129-34.

Pereira, M. C., & Lopes, A. S. (2014). Time-varying fiscal policy in the US. *Studies in Nonlinear Dynamics & Econometrics*, *18(2)*, 157-184.

Ramey, V. A. (2011). Identifying government spending shocks: It's all in the timing. *The Quarterly Journal of Economics*, *126*(1), 1-50.

Ramey, V. A., & Zubairy, S. (2018). Government spending multipliers in good times and in bad: evidence from US historical data. *Journal of Political Economy*, *126*(2), 850-901.

Rothfelder, M., & Boldea, O. (2019). Testing for a threshold in models with endogenous regressors. (CentER Discussion Paper; Vol. 2019-030)

Stock, J. H., & Watson, M. W. (2018). Identification and estimation of dynamic causal effects in macroeconomics using external instruments. *The Economic Journal*, *128*(610), 917-948.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, *58*(1), 267-288.

Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American statistical association*, *101*(476), 1418-1429.

Received: Nov 09, 2022 Revised: Nov 15, 2022 Accepted: Dec 30, 2022