Journal of Applied Econometrics

WILEY

Journal of APPLIED ECONOMETRICS

REPLICATION OPEN ACCESS

Reassessing the Cross-Sectional Fiscal Multiplier: Evidence From U.S. Defense Procurement, 1966–2019

Gianluca Pallante

Institute of Economics and l'EMbeDS, Sant'Anna School of Advanced Studies, Pisa, Italy

Correspondence: Gianluca Pallante (gianluca.pallante@santannapisa.it)

Received: 23 April 2025 | **Revised:** 31 January 2026 | **Accepted:** 2 March 2026

Keywords: cross-sectional multipliers | defense procurement | fiscal policy | government expenditures

ABSTRACT

This paper revisits the empirical analysis of Nakamura and Steinsson (2014). I reconstructed and extended the original dataset to cover the period 1966–2019, harmonizing two major sources of data: the Defense Contract Action Data System (DCADS) and USAspending.gov. I discuss how to aggregate these contract-level data to better capture spending more directly tied to domestic stimulus. Estimated multipliers are slightly lower in narrow replications but increase when incorporating later fiscal episodes. I also assess the validity and stability of cross-sectional estimates. While some heterogeneity exists, dispersion in state-level responses remains within reasonable boundaries, especially when accounting for dynamic persistence.

JEL Classification: E62, H57, C18, R12

1 | Introduction

The role of fiscal policy in macroeconomic stabilization has gained renewed prominence following the 2007–2009 financial crisis, particularly as monetary policy faced constraints at the zero lower bound (Ramey 2011; Blanchard et al. 2012). A growing body of research has examined the effects of fiscal policy interventions (Ramey 2019), highlighting their asymmetry in expansions and consolidations (Blanchard and Leigh 2013; Barnichon et al. 2022), state-dependency (Auerbach and Gorodnichenko 2012; Ferraresi et al. 2015), and variation across policy regimes (Ilzetzki et al. 2013).[1] Beyond its role in short-term stabilization, fiscal policy has also been instrumental in driving long-term economic transformations. Large-scale government spending, especially in defense and research and development (R&D), has historically spurred industrial growth and technological advancements (Pallante et al. 2023; Gross and Sampat 2023; Moretti et al. 2025; Kantor and Whalley 2025). Empirical evidence further suggests that government spending can have persistent effects on innovation, productivity, and private sector investment (Antolin-Diaz and Surico 2025; Fieldhouse and Mertens 2024; Ilzetzki 2024), reinforcing its relevance for long-run economic performance.

Given the importance of fiscal interventions, measuring their impact remains a key empirical challenge. The fiscal multiplier, defined as the change in output per dollar of government spending, has been traditionally been estimated using time-series approaches. While useful, these methods face limitations: large exogenous spending shocks are rare, and disentangling their effects from concurrent policy changes can be difficult (Nakamura and Steinsson 2018). An alternative approach, increasingly used in empirical macroeconomics, leverages cross-sectional variation in government spending across smaller geographies. By comparing regions that experience different levels of spending shocks, cross-sectional studies can overcome some of the limitations of time-series models, providing more detailed insights on transmission mechanisms and distributional issues (Chodorow-Reich 2019).

A seminal contribution in this literature is Nakamura and Steinsson (2014, NS14, henceforth), which estimates local fiscal multipliers using variation in U.S. military procurement across states. Their identification strategy assumes that changes in aggregate military spending are exogenous to local economic conditions, while its allocation across states follows historical procurement patterns. This approach provides a quasi-experimental setting to estimate how regional economies respond to government demand shocks. Their findings suggest a multiplier of approximately 1.5. NS14 has become a benchmark study, stimulating debates on whether or not there is "an applied micro free lunch for macroeconomists" (Ramey 2019).

Indeed, studies suggest that cross-sectional multipliers may be sensitive to heterogeneous regional responses (Andrews 2005; Pesaran 2006; Almuzara and Sancibrián 2024). If geographical units respond differently to policy shocks, standard cross-sectional methods may produce unstable or biased estimates, even when adopting a valid research design.

This paper extends and assesses the robustness of NS14's empirical analysis by reconstructing their dataset and expanding it to the period 1966–2019, including 13 more years, and thus capturing additional waves of military spending expansions and contractions and covering a period where the U.S. monetary policy regime is constrained by the zero lower bound. One contribution is the harmonization of military procurement data from two distinct sources: the Defense Contract Action Data System (DCADS), used in NS14, and USAspending.gov, which reports post-2006 U.S. government expenditures, including those appropriated by the Department of Defense. To provide further reliability of cross-sectional estimates, a methodological contribution of this paper is to assess the stability of cross-sectional multiplier estimates using the diagnostic proposed by Canova (2024) and implemented in Canova and Pappa (2025). This procedure tests whether dispersion in the location-specific estimated multipliers is sufficiently low to validate the cross-sectional approach.

The findings confirm that fiscal multipliers remain positive, statistically different from zero and greater than one. However, they are lower when narrowly compared with NS14. Yet, they increase when the post-2006 period is included in the analysis. The state-level aggregation that matches the NS14 estimates the closest and most plausibly isolates deficit-financed government stimulus excludes contracts designed for foreign military sales and standardizes the minimum threshold reporting requirement across the whole sample. Finally, using Canova (2024)'s proposed test, results indicate that when accounting for dynamic persistence, the state-level cross-sectional multipliers are relatively stable and heterogeneity in the dynamic response is not statistically detected, validating the cross-sectional approach.

The paper is structured as follows: Section 2 details how the defense-related spending series is constructed and aggregated at the state level. Section 3 recaps the framework and the research design in NS14. Sections 4 and 5 replicate and test the stability of the cross-sectional multipliers, respectively. Section 6 concludes.

2 | Data and Aggregation Filters

2.1 | Data Description

The empirical analysis relies on two primary data sources to construct a measure of defense-related spending across U.S. states. The first is the Defense Contract Action Data System (DCADS). This is the Department of Defense's data collection system for reporting contract actions to the Federal Procurement Data System (FPDS), which is the authoritative source of contract information for all U.S. government agencies. DCADS reports prime contracts for goods and services between the private sector and agencies of the Department of Defense. In particular, each contract record relies on the DD 350 form, that is, the Individual Contracting Action Report, including detailed information on the contracting office, recipient firm, place of performance, whether it is targeted for a foreign sale or not, and the total contract value.[2] The second source of data is USAspending.gov, the official open-data repository that provides records about contracts awarded by the U.S. government, including contracts, grants, and financial assistance programs. I select information on prime contracts from the Department of Defense. This source provides more detailed information than DCADS, but it can be matched with the latter along dimensions such as location, dates, contractor and contract characteristics.

Following NS14, the military spending dataset is complemented with additional macroeconomic variables, including measures of state output, sectoral output, population, inflation, employment, and military compensation.[3] The final dataset covers 51 U.S. states, including D.C., and spans the period 1966–2019, thus incorporating 13 additional years of defense contract data.

2.2 | State-Level Aggregation

After collecting and merging the contract-level data from DCADS and USAspending.gov, the dataset comprises approximately 41.1 million individual contract observations. Since NS14 estimates state-level fiscal multipliers, the contract-level data must be aggregated at the state-year level to construct a comparable measure of defense spending. I first perform the easiest form of aggregation by summing all observations grouped by the 51 state names and year. I label this dataset as "RAW," as it provides the unadjusted measure of spending. Given the characteristics of the two data sources assembled, to ascertain the role played by measurement errors, I apply three different data filters, briefly summarized in Table 1 for convenience.

Threshold Reporting Requirements (T25). From 1966 to 1983, the records include contracts with a value of $10,000 or larger. From fiscal year 1983 onward, the records include contracts with a value of $25,000 or more. However, USAspending.gov comprises transactions lower than these reporting requirements. Many of these transactions include but are not limited to revisions (see also Auerbach et al. 2020, for a discussion). For this reason, all contracts that report amounts less than $25,0000 in absolute value are removed. Although the authors acknowledge changes in the threshold for reporting, I document that excluding contracts below the 25k threshold may avoid spurious correlation

TABLE 1 | Summary of state-level aggregation methods.

Method	Description
A: RAW	Baseline aggregation: all contract-level transactions are summed at the state-year level without further adjustments.
B: GEO (Geographical adjustment)	Corrects inconsistencies in state identification (misspellings, miscoding, missing state labels) using additional location information. Also removes inconsistencies due to deobligations associated with repeated contract entries.
F: T25 (Threshold filter)	Excludes contracts reporting an obligation amount below \$25,000 in absolute value. Addresses inconsistencies in federal reporting thresholds.
G: FMS (Foreign Military Sales exclusion)	Removes transactions financed or co-financed by foreign governments or international organizations.

Note: Combined methods (e.g., C: GEO+T25, D: GEO+FMS, E: GEO+T25+FMS, H: T25+FMS) apply the corresponding filters jointly.

due to the possible concentration of small or large contracts in particular states.

Geographical Adjustment (GEO). The dataset shows inconsistencies in state name coding, arising from misspellings, incorrect abbreviations, or missing entries. To assign a geographical reference to these observations, I manually retrieve the correct state name based on county and place of performance names or identifiers. Moreover, the USAspending.gov dataset contains many contracts awarded to the same contractor, performed in the same place, with deobligated amounts very close (or identical) to the original (first-awarded) one. Similar to Demyanyk et al. (2019), both contracts are removed from the dataset whenever the difference between the original contract and the corresponding deobligation is lower than the 0.5% of deobligated amount. This filter eliminates about 4.4% of the observations (very close to Demyanyk et al. (2019)), which are concentrated in the post-2008 period. To ensure consistency, I apply the same filter to the DCADS records, where such instances do not occur. By correcting these inconsistencies, this filter lowers the measurement error in the geographical distribution of spending across states.

Foreign Funding for Military Acquisitions (FMS). There are cases in which a foreign government, international organization, or foreign military organizations bear some of the cost of the acquisition. The Foreign Military Sale (FMS) is a U.S. program for transferring military equipment or services to international partners.[4] Since FMS contracts are not financed through domestic U.S. government spending, they do not represent true fiscal stimulus within the U.S. economy. Including FMS transactions in the dataset would therefore introduce measurement error in the estimation of fiscal multipliers. Consequently, when FMS-related transactions are excluded, the state-level aggregate of military contracts reflects only domestically financed spending. In some phases of the U.S. foreign policy, contracts under FMS were not negligible and therefore I suggest that removing this type of spending may help avoid errors in the measurement of a deficit-financed public spending variable.

Figure 1 shows the discrepancies when accounting for different methods of data adjustments. I acknowledge that none of the filters exactly mirror the data published along NS14. Filter A (No adjustments in the geographical imputations, i.e, RAW) deviates quite a lot from the dataset I aim to replicate, with a right-skewed discrepancy that averages 4.7%. On the contrary, filter D (GEO and FMS adjustments) matches NS14 data the closest, with a discrepancy that is equally distributed around zero. Thus, in order to achieve the closest replication, the researcher should exclude contracts stipulated for foreign sales and should not neglect inconsistencies in the geographical accounting of the entries. To provide further detail, Table 2 summarizes where these discrepancies are mostly concentrated. The largest positive differences are observed in a small number of states (Kentucky, DC, and Alabama, accounting together for roughly 5% of DoD spending), where average gaps reach roughly 10%–30% depending on the aggregation filter. By contrast, the largest negative discrepacies are generally smaller in absolute magnitude and more sensitive to the filter applied; with the exception of states such as Missouri and Texas—which together receive about 11% of total federal DoD spending and exhibit gaps on the order of 8%–10%—the remaining bottom-tail states, collectively accounting for about 1.2% of DoD spending, display comparatively modest deviations. Along the time dimension, positive discrepancies are more concentrated in the later part of the sample. Overall, these patterns indicate that discrepancies are not uniform across states or over time.[5] To provide more context, the battery of estimations that I perform clarify whether these aggregation procedures affect or not substantially the size of the fiscal multiplier.

3 | Model & Research Design

To estimate the fiscal multiplier, I adopt the NS14 regression specification, which is also widely used in the estimation of local fiscal multipliers (Auerbach et al. 2020, 2024; Muratori et al. 2023):

$$\frac{Y_{i,t} - Y_{i,t-2}}{Y_{i,t-2}} = \beta \frac{G_{i,t} - G_{i,t-2}}{Y_{i,t-2}} + \alpha_i + \gamma_t + \varepsilon_{i,t} \tag{1}$$

Closely following NS14, $Y_{i,t}$ measures the real per capita output in state i in year t, while $G_{i,t}$ captures the real per capita value of public spending, measured by Department of Defense (DoD) procurement contracts. The specification includes state fixed effects α_i to account for state-specific trends in output and military spending, as well as year fixed effects γ_t to control for aggregate economic fluctuations and policies carried out at the national level. The coefficient of interest, β, represents the fiscal multiplier, which quantifies the dollar value of output generated by a one-dollar increase in government expenditures. Standard errors are clustered at the state level to account for serial correlation induced by the overlapping observation windows.

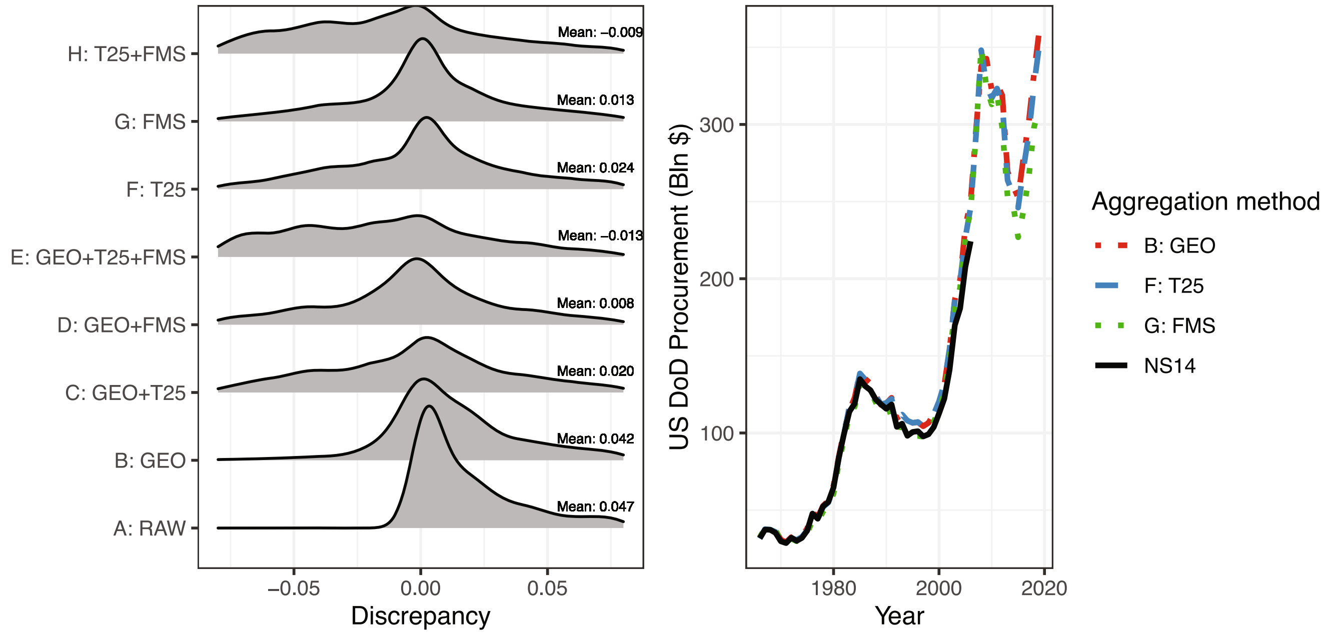

FIGURE 1 | Differences with original data from NS14 and with different state-level aggregation filters. Left panel: density plots of percentage difference between state-year observations in NS14 and using the data filtering procedures described in Section 2. Right panel: US DoD procurement series in NS14 and for selected filtering methods. See Table 1 for a brief description of the state-level aggregation methods.

TABLE 2 | Largest discrepancies in the reconstructed spending series by state and year across the aggregation methods proposed.

	I. states						II. fiscal years					
Aggregation	**Top 3**			**Bottom 3**			**Top 3**			**Bottom 3**		
A: RAW	KY	DC	AL	ME	RI	MT	2006	2005	1994	1976	1989	1975
	(0.38)	(0.19)	(0.13)	(0.01)	(0.01)	(−0.03)	(0.14)	(0.13)	(0.12)	(0.02)	(0)	(−0.01)
B: GEO	KY	DC	AL	GA	RI	MT	2006	2005	1994	1981	1989	1975
	(0.37)	(0.19)	(0.13)	(0)	(0)	(−0.03)	(0.14)	(0.13)	(0.11)	(0)	(0)	(−0.02)
F: T25	KY	DC	AL	ID	OR	MT	1994	2006	2002	1976	1972	1975
	(0.32)	(0.17)	(0.11)	(−0.03)	(−0.03)	(−0.05)	(0.12)	(0.11)	(0.1)	(−0.03)	(−0.04)	(−0.05)
G: FMS	KY	DC	AL	VT	MO	TX	2006	2005	2004	1977	1976	1975
	(0.38)	(0.19)	(0.12)	(−0.07)	(−0.08)	(−0.09)	(0.12)	(0.1)	(0.09)	(−0.04)	(−0.05)	(−0.07)

Note: For the three aggregation methods, the Table reports the three states (Panel I) and three fiscal years (Panel II) with the largest positive (Top 3) and largest negative (Bottom 3) average discrepancies, computed as the mean percentage gap relative to the NS14 state-level procurement series. Table S2 in the online Supporting Information provides the discrepancies for all filtering combinations. See Table 1 for definitions of the aggregation methods.

A central challenge in estimating the fiscal multiplier is the endogeneity of government spending, as defense procurement may be influenced by state-specific economic conditions. Political and strategic factors often play a role in allocating defense contracts, either by directing spending to struggling regions as an implicit form of economic support or by prioritizing politically influential states. If unobserved economic factors influence both government spending and economic performance, OLS estimates of β will be biased.

To address this concern, NS14 adopts an instrumental variable (IV) strategy, leveraging the fact that state-level military spending responds systematically in a different way to fluctuations in the total U.S. defense budget. The identification assumption is that while total military spending is driven by national security considerations and geopolitical events, its state-level allocation also follows historical procurement patterns. This allows the isolation of the systematic response of state-level expenditures to total U.S. military spending.

The two-stage least squares interpretation of the estimated coefficient is that only systematic state-level variation in military spending is predicted. The first-stage equation is $\Delta G_{i,t} = \theta_i \Delta G_t + \delta_i + \lambda_t + \nu_{i,t}$, where G_t represents total U.S. military spending, and θ_i captures state-specific exposure to national defense appropriations. This exposure coefficient reflects the systematic tendency of state i to receive military contracts when the federal defense budget changes. Throughout the text, only 2SLS regression results are reported.

4 | Results

Do the filters that more closely match the empirical distribution of the original NS14 data deliver the same estimates of the fiscal multiplier? Figure 3 presents a summary of the estimated multipliers under different data aggregation approaches, providing a comparison across alternative specifications.[6] To broaden the scope of the replication study, all estimations are performed using the R software, with the `fixest` package. The figure contrasts the narrowest replication (dashed lines), retrieving the original data from NS14 and covering the 1966–2006 period, with estimates obtained by incorporating data up to 2019. To assess the influence of measurement error in complementary variables, which may be due to revisions, I report estimates (dotted lines) based on a dataset assembled using the original NS14 spending series and retrieving all additional variables from their primary sources, as documented in Table S1 in the online Supporting Information. These estimates serve as a benchmark against those obtained using the replicated, harmonized, and extended dataset, which varies depending on the filtering applied. As in NS14, all output multipliers are reported both including and excluding military compensation from the spending series (in red circles and blue triangles, respectively).

Across all specifications, the estimates are very close to those reported in NS14, systematically lower, yet greater than one. In particular, the NS14 benchmark estimate is typically contained within the confidence interval around the corresponding replicated estimates.[7] Once the sample is extended to 2019 and incorporates the harmonized procurement series, the estimated multipliers increase and get closer to the original NS14 benchmark (right-hand panels of Figure 3). This finding is somewhat consistent with the interpretation that the effects of government spending are more prominent during periods of economic slack or when monetary policy hits the zero lower bound, an economic regime effectively covered by the post-2006 extended sample.

Interestingly, the aggregation method that minimizes the average discrepancy with NS14's data—filter D (GEO+FMS)—does not necessarily yield the closest point estimates of the fiscal multiplier. Nonetheless, excluding contracts flagged as foreign military sales (FMS) appears to have the most pronounced effect in raising the estimated multiplier. This effect is followed, in magnitude, by the implementation of a $25,000 threshold on contract reporting (T25). Improving the geographical allocation of contracts (GEO) seems to exert a downward pressure on the estimates, revealing the presence of a mild attenuation bias. Among these, the FMS filter is of particular interest as it isolates procurement more directly attributable to domestic demand stimulus and likely financed through deficit spending. Given its theoretical and empirical implications, results suggest that this specification is the preferred one. This is also supported by Figure 2, which plots the two-year changes in output against the predicted two-year changes in DoD procurement, across the main aggregation filters, distinguishing observations in the pre- and post-2006 subsamples. Across all filters, enough of the overall variation in spending occurring in the post-2006 period (cf., Figure 1, panel B) still contributes to a positive IV coefficient. Particularly, the exclusion of FMS contracts helps strengthen this relationship by aligning better with the results obtained with the NS14 sample and the pre-2006 period.

The estimated multipliers also replicate one of the key empirical regularities reported in NS14: the size of the fiscal multiplier increases with the geographic scope of aggregation. The lower panels of Figure 3 show that estimates computed at the Census region level are consistently larger than those at the state level. Additionally, excluding military compensation from the spending series reduces the estimated multiplier at the state level but increases it at the regional level, a pattern also observed in NS14.[8]

Employment and sectoral results, as reported in the online Supporting Information, broadly confirm the patterns found for output. Employment multipliers are typically below the NS14 estimates, with the exclusion of contracts associated with foreign military sales (FMS) again playing an important role in bringing estimates closer to the NS14 benchmark. Relative to output, employment and sectoral responses are less sensitive to alternative aggregation choices. In particular, the latter are statistically different from zero in construction, manufacturing, retail, and services and tend to be smaller when the sample is extended through 2019.[9]

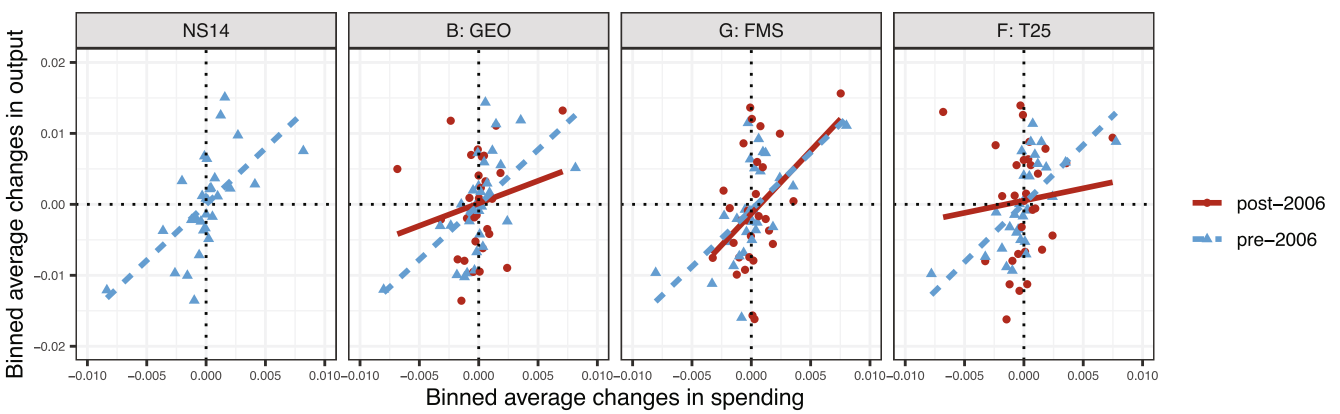

FIGURE 2 | Quantiles of change in output versus predicted change in military spending. The Figure shows, for the original NS14 sample and across the main aggregations (see Table 1 for definitions), averages of two-year changes in output against predicted two-year changes in DoD procurement (based on the first-stage regression). As in NS14, averages are computed by grouping observations into 30 quantiles of the predicted spending distribution, both for the pre- and post-2006 period (triangles and dots, respectively). The Figure also shows the slope coefficient from the binned regression fit for the two subsamples (dashed for pre-2006 and solid for the post-2006).

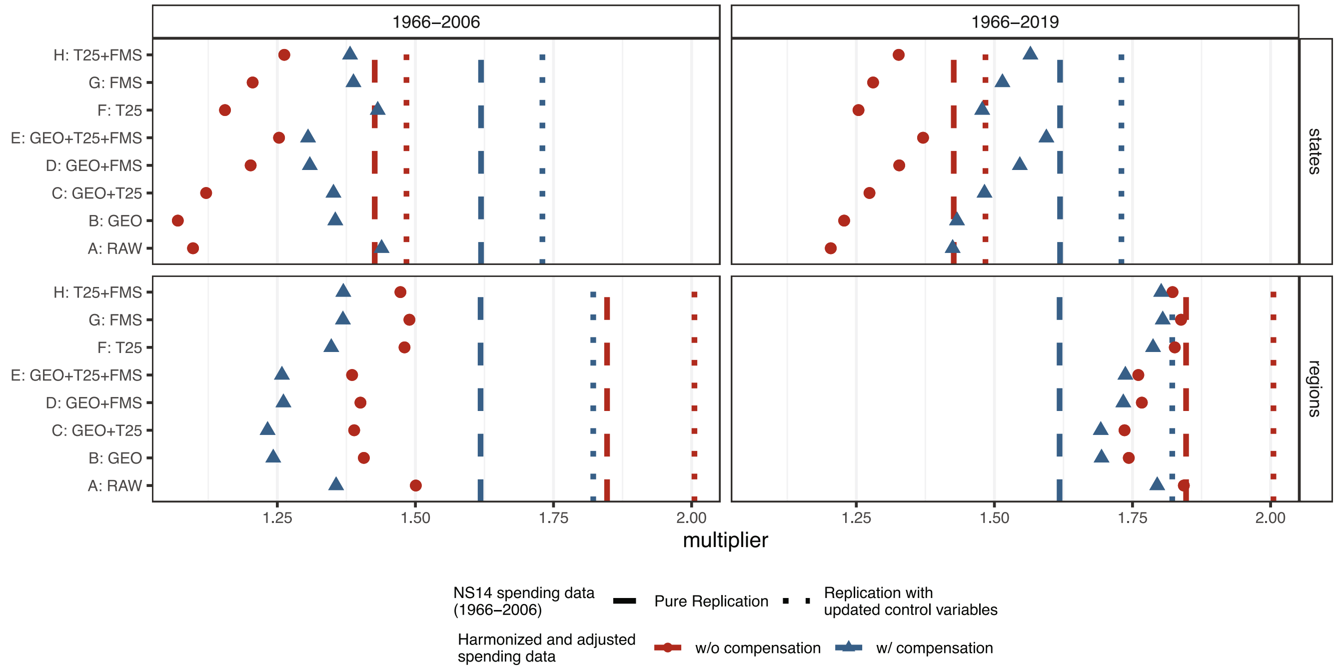

FIGURE 3 | Revisited estimates of cross-sectional fiscal spending output multipliers. The Figure reports 2SLS estimates of the output multiplier from Equation (1), computed at the state and Census-region levels, for the original sample (1966–2006) and the extended sample (1966–2019), and for each aggregation filter (A–H, see Table 1 for definitions). For the NS14 spending series, estimates are shown under (i) *Pure replication* (dashed lines) and (ii) *Replication with updated controls* (dotted lines). Estimates under *Harmonized and adjusted spending data*, include both retrieved controls and the extended spending series. The figures distinguish multipliers estimated when the spending series includes DoD compensations (blue triangles) or not (red dots).

5 | Testing Dynamic Heterogeneity

Recent methodological contributions have raised concerns about the reliability of cross-sectional macroeconomic estimates, particularly in settings where policy interventions induce heterogeneous dynamic responses across units. If regions or states respond differently to the same policy shock, cross-sectional estimation methods may fail to produce stable and reliable multiplier estimates.

To evaluate the robustness of the U.S. defense-related fiscal multiplier, I apply the diagnostic test proposed by Canova (2024), which relies on the coefficient of variation (CV) of the distribution of region-specific estimated multipliers. The underlying rationale is that if responses to policy shocks are homogeneous across states, the CV should be small, indicating that estimates are stable. Conversely, if responses are highly heterogeneous, the CV will be large, signaling that cross-sectional methods may not provide an accurate representation of the underlying economic effects. The diagnostic relies on computing the distribution of *state-by-state* estimated multipliers and comparing the CV with bootstrap-generated critical values used to assess the null hypothesis of homogeneous regional response. The degree to which estimates vary across states serves as an indicator of the extent of heterogeneity in fiscal policy responses.

Since the number of regional aggregates (10) is too small to conduct a meaningful CV test, I restrict the analysis to the state-level dataset, which consists of 51 states over 53 years. I adopt two M specifications commonly used in the estimation of fiscal multipliers. The first specification (M = NS) estimates the policy effect using the standard NS14 approach, while the second specification (M = LagDep) introduces lagged dependent variables to account for heterogeneity in dynamic persistence. In both specifications, the policy variable enters in first differences and the *state-by-state* estimate of the two-year multiplier is $\beta_{i,M}$.[10] More formally:

$$\text{(NS)} \quad \frac{Y_{i,t} - Y_{i,t-2}}{Y_{i,t-2}} = \beta_{i,NS} \frac{G_{i,t} - G_{i,t-2}}{Y_{i,t-2}} + \alpha_i + \varepsilon_{i,t} \tag{2}$$

$$\text{(LagDep)} \quad \frac{Y_{i,t} - Y_{i,t-2}}{Y_{i,t-2}} = \beta_{i,LagDep} \frac{G_{i,t} - G_{i,t-2}}{Y_{i,t-2}} + \delta_{1,i} \frac{Y_{i,t-1} - Y_{i,t-3}}{Y_{i,t-3}} + \delta_{2,i} \frac{G_{i,t-1} - G_{i,t-3}}{Y_{i,t-3}} + \alpha_i + \varepsilon_{i,t} \tag{3}$$

Importantly, in both models, the multiplier and the dynamic coefficients are allowed to be unit-specific: heterogeneity may arise in the impact response, in propagation, or in both. In particular for the LagDep model, $\delta_{1,i}$ and $\delta_{2,i}$ capture such heterogeneous persistence and explicitly accounting for it can then reduce cross-sectional dispersion in the multiplier of interest that may reflect differences in propagation. Thus, Equation (3) helps distinguish heterogeneity in propagation from heterogeneity in the instantaneous policy effect. Results in Table 3 focus on the contemporaneous multipliers. The NS specification exhibits greater dispersion, with higher CV values compared to the LagDep specification. This suggests that the assumption of homogeneous responses across states is less defensible when

TABLE 3 | Coefficient of variation diagnostic for heterogeneous dynamic responses.

	NS			LagDep				NS			LagDep		
	β	$\overline{\beta}$	$\lvert IQR\rvert$	β	$\overline{\beta}$	$\lvert IQR\rvert$		β	$\overline{\beta}$	$\lvert IQR\rvert$	β	$\overline{\beta}$	$\lvert IQR\rvert$
A: RAW	1.10	−1.04	2.63	1.37	2.95	0.90	E: GEO+T25+FMS	1.25	−0.95	4.83	1.49	3.30	0.94
B: GEO	1.07	−0.83	2.77	1.40	3.20	0.77	F: T25	1.16	−1.08	2.75	1.45	3.09	0.91
C: GEO+T25	1.12	−0.89	3.00	1.48	3.34	0.84	G: FMS	1.21	−1.05	3.41	1.37	3.00	1.02
D: GEO+FMS	1.20	−0.89	3.78	1.42	3.16	0.94	H: T25+FMS	1.26	−1.12	3.91	1.45	3.12	1.00
|IQR| Critical values				(2-year ahead multiplier)									
	90% = 1.425		0.95% = 1.576		0.99% = 1.903								

Note: The table report, for each data filter and model specification as in Equations (2) and (3), the cross-sectional aggregate estimate β, the average multiplier computed across state-by-state regressions $\overline{\beta}$, and the coefficient of variation defined as the absolute value of the interquartile range of the estimated distribution of multipliers divided by the midpoint of the interquartile range, as in Canova (2024). The critical values for testing the null hypothesis of homogeneity in the dynamic response are taken from Canova (2024), Table 3. See Table 1 for a brief description of the state-level aggregation methods.

lagged economic conditions are not controlled for explicitly. The corresponding critical values for the null of homogeneity are 1.425 (90%), 1.576 (95%), and 1.903 (99%) for the two-year-ahead multiplier. The estimated CV values for the LagDep specification remain lower than the reported critical values, suggesting that controlling for the persistence in economic conditions is important for taming the heterogeneity in state-level responses. While some degree of variation across states is observed, the results indicate that state-level heterogeneity does not fundamentally undermine the cross-sectional approach. However, the differences between the NS and LagDep specifications suggest that modeling dynamic persistence explicitly can reduce dispersion in multiplier estimates, making the latter specification preferable.

6 | Conclusion

This paper revisits the cross-sectional estimates of the U.S. defense-related fiscal multiplier, retrieving and extending the empirical exercise of Nakamura and Steinsson (2014, NS14). Data from the Defense Contract Action Data System (DCADS) and USAspending.gov have been harmonized to build a consistent series of defense-related expenditures across U.S. states from 1966 to 2019.

The analysis shows that, when using this detailed source of data, the estimation of the cross-sectional multiplier is influenced by changes in the minimum threshold for procurement reporting, geographical assignment of performance and whether the signed contracts are designated for foreign military sales. In particular, researchers are advised to exclude contracts with a foreign sale destination as this adjustment brings the empirical estimates closer to the NS14 benchmark, but it also better captures spending more directly attributable to a domestic, deficit-financed demand stimulus.

Indeed, the results indicate that the fiscal multipliers estimated using the harmonized and extended dataset remain broadly consistent with those originally reported by NS14. While the narrow replication yields slightly lower estimates, extending the dataset to include post-2006 observations leads to larger multipliers—consistent with the literature emphasizing stronger fiscal effects during periods of economic slack and where monetary policy is also stuck at the zero lower bound.

Furthermore, challenged by recent methodological contributions suggesting the conditions under which cross-sectional macro elasticities are unstable in the presence of heterogeneous dynamic response, I test the stability of the NS14 empirical framework using a coefficient of variation (CV) diagnostic recently proposed by Canova (2024). The analysis shows that, although some heterogeneity exists across states, the dispersion of estimates generally remains within the critical bounds. The inclusion of lagged dependent variables reduces variability and enhances the stability of the estimates, reinforcing the importance of accounting for dynamic persistence. These findings confirm that cross-sectional methods, when properly implemented and tested, continue to provide a valid empirical framework for evaluating the macroeconomic effects of fiscal interventions.

Acknowledgements

I wish to warmly thank Damiano di Francesco for his comments and interesting discussions. Finally, I am grateful to the editor Heather Anderson and the three reviewers for their comments and guidance throughout the process. Open access publishing facilitated by Scuola Superiore Sant'Anna, as part of the Wiley - CRUI-CARE agreement.

Funding

I acknowledge financial support under the National Recovery and Resilience Plan (NRRP), Mission 4, Component 2, Investment 1.1, funded by the European Union - NextGenerationEU- Project Title "Steering finance to Wards green technologies: a hybrid experimental-computational macro approach", P2022FEY25 - CUP 153D23016680001 by the Italian Ministry of Ministry of University and Research (MUR).

Data Availability Statement

The data that support the findings of this study are openly available in the Journal Data Archive at https://doi.org/10.15456/jae.2026091.1425247126.

Endnotes

[1]The list is far from being exhaustive. For a more recent review of the major empirical contributions, see Castelnuovo and Lim (2019).

[2]The DCADS archival records can be found in two different series: "Records of Prime Contracts Awarded by the Military Services and Agencies", (ARC Identifier 606901) for the period 1965–1975 and

available at https://catalog.archives.gov/id/606901; for fiscal years 1976–2006, instead I use the "Records of Prime Contracts Awarded by the Military Services and Agencies" (ARC Identifier 578589), available at https://catalog.archives.gov/id/578589. These series differ in record layout across fiscal years; I therefore rely on the year-specific archival documentation to select the attributes of interest. Since prime contract microdata is no longer accessible via the same link used in NS14, reconstructing the series from these sources may lead to minor discrepancies relative to the published NS14 aggregates.

[3]See Table S1 in the online Supporting Information online to retrieve the economic indicators and their sources.

[4]More info on the FMS program at https://www.dsca.mil/foreign-military-sales-fms.

[5]Figure S1 in the online Supporting Information shows how the full cross-sectional and time-series distribution changes across the aggregation filters.

[6]For brevity and clarity, I limit the focus of the paper's findings to a subset of all NS14's results, which include a series of robustness checks such as the use of a classic Bartik instrument, the estimation of the fiscal multipliers under alternative specifications and in periods of high versus low unemployment. The full set of replication results are reported in Tables S3 and S4 in the online Supporting Information. For transparency, Table S8 in the online Supporting Information reports the baseline estimates obtained using aggregates constructed with the original Stata aggregation routine, kindly shared by Emi Nakamura; instructions to run the corresponding R script are also provided in the Supporting Information.

[7]The corresponding confidence intervals for the estimates of output and employment multipliers are reported in the online Supporting Information, Figure S2. For transparency, in all Tables the document also highlights in bold those cases in which the NS14 point estimate falls outside the confidence interval implied by my estimates (though, in practice, this does not occur in the specifications reported).

[8]As a robustness check, I estimate the baseline state-level output multiplier by dropping states or years that exhibit significant discrepancies, as reported in Table 2, and sensitivities $\hat{\theta}_i$ to changes in federal defense spending (See first-stage discussion in Section 3). Across specifications and aggregation filters, estimates remain similar to the baseline results without evidence of state and years acting as outliers. As in NS14, excluding more (less) sensitive states slightly raises (lowers) the estimated multiplier. Full details are reported in Table S7 in the online Supporting Information.

[9]Sectoral output is measured using SIC classifications prior to 1997 and NAICS classifications thereafter. Because 1997 is available under both systems, growth rates can be linked consistently. Macro-regional sectoral estimates are reported in the online Supporting Information, Tables S5 and S6.

[10]To make the two approaches more easily comparable, only the contemporaneous two-year difference in $G_{i,t}$ is instrumented as detailed in Section 3.

References

Almuzara, M., and V. Sancibrián. 2024. *Micro Responses to Macro Shocks* (Staff Report No. 1090). New York, NY: Federal Reserve Bank of New York. https://doi.org/10.59576/sr.1090.

Andrews, D. W. K. 2005. "Cross-Section Regression With Common Shocks." *Econometrica* 73, no. 5: 1551–1585. https://doi.org/10.1111/j.1468-0262.2005.00629.x.

Antolin-Diaz, J., and P. Surico. 2025. "The Long-Run Effects of Government Spending." *American Economic Review* 115, no. 7: 2376–2413. https://doi.org/10.1257/aer.20231278.

Auerbach, A., Y. Gorodnichenko, and D. Murphy. 2020. "Local Fiscal Multipliers and Fiscal Spillovers in the USA." *IMF Economic Review* 68, no. 1: 195–229. https://doi.org/10.1057/s41308-019-00102-3.

Auerbach, A. J., and Y. Gorodnichenko. 2012. "Measuring the Output Responses to Fiscal Policy." *American Economic Journal: Economic Policy* 4, no. 2: 1–27. https://doi.org/10.1257/pol.4.2.1.

Auerbach, A. J., Y. Gorodnichenko, and D. Murphy. 2024. "Macroeconomic Frameworks: Reconciling Evidence and Model Predictions From Demand Shocks." *American Economic Journal: Macroeconomics* 16, no. 3: 190–229. https://doi.org/10.1257/mac.20210033.

Barnichon, R., D. Debortoli, and C. Matthes. 2022. "Understanding the Size of the Government Spend-Ing Multiplier: It's in the Sign." *Review of Economic Studies* 89, no. 1: 87–117. https://doi.org/10.1093/restud/rdab029.

Blanchard, O. J., and D. Leigh. 2013. "Growth Forecast Errors and Fiscal Multipliers." *American Economic Review* 103, no. 3: 117–120. https://doi.org/10.1257/aer.103.3.117.

Blanchard, O. J., D. Romer, A. M. Spence, and J. E. Stiglitz, eds. 2012. *In the Wake of the Crisis: Leading Economists Reassess Economic Policy*. MIT Press. https://www.jstor.org/stable/j.ctt5vjq9h.

Canova, F. 2024. "Should We Trust Cross-Sectional Multiplier Estimates?" *Journal of Applied Econometrics* 39, no. 4: 589–606. https://doi.org/10.1002/jae.3041.

Canova, F., and E. Pappa. 2025. "The Macroeconomic Effects of EU Regional Structural Funds." *Journal of the European Economic Association* 23, no. 1: 327–360. https://doi.org/10.1093/jeea/jvae031.

Castelnuovo, E., and G. Lim. 2019. "What Do we Know About the Macroeconomic Effects of Fiscal Policy? A Brief Survey of the Literature on Fiscal Multipliers." *Australian Economic Review* 52, no. 1: 78–93. https://doi.org/10.1111/1467-8462.12313.

Chodorow-Reich, G. 2019. "Geographic Cross-Sectional Fiscal Spending Multipliers: What Have we Learned?" *American Economic Journal: Economic Policy* 11, no. 2: 1–34. https://doi.org/10.1257/pol.20160465.

Demyanyk, Y., E. Loutskina, and D. Murphy. 2019. "Fiscal Stimulus and Consumer Debt." *Review of Economics and Statistics* 101, no. 4: 728–741. https://doi.org/10.1162/rest_a_00796.

Ferraresi, T., A. Roventini, and G. Fagiolo. 2015. "Fiscal Policies and Credit Regimes: A TVAR Approach." *Journal of Applied Econometrics* 30, no. 7: 1047–1072. https://doi.org/10.1002/jae.2420.

Fieldhouse, A. J., and K. Mertens. 2024. *The Returns to Government R&D: Evidence From U.S. Appropriations Shocks* (Working Paper No. 2305). Dallas, TX: Federal Reserve Bank of Dallas. https://doi.org/10.24149/wp2305r2.

Gross, D. P., and B. N. Sampat. 2023. "America, Jump-Started: World War II R&D and the Takeoff of the US Innovation System." *American Economic Review* 113, no. 12: 3323–3356. https://doi.org/10.1257/aer.20221365.

Ilzetzki, E. 2024. "Learning by Necessity: Government Demand, Capacity Constraints, and Produc-Tivity Growth." *American Economic Review* 114, no. 8: 2436–2471. https://doi.org/10.1257/aer.20230033.

Ilzetzki, E., E. G. Mendoza, and C. A. Végh. 2013. "How Big (Small?) Are Fiscal Multipliers?" *Journal of Monetary Economics* 60, no. 2: 239–254. https://doi.org/10.1016/j.jmoneco.2012.10.011.

Kantor, S., and A. Whalley. 2025. "Moonshot: Public R&D and Growth." *American Economic Review* 115, no. 9: 2891–2925. https://doi.org/10.1257/aer.20220540.

Moretti, E., C. Steinwender, and J. Van Reenen. 2025. "The Intellectual Spoils of War? Defense R&D, Productivity, and International Spillovers." *Review of Economics and Statistics* 107, no. 1: 14–27. https://doi.org/10.1162/rest_a_01293.

Muratori, U., P. Juarros, and D. Valderrama. 2023. *Heterogeneous Spending, Heterogeneous Multipliers* (IMF Working Paper No. 23/52). Washington, DC: International Monetary Fund. https://doi.org/10.5089/9798400233777.001.

Nakamura, E., and J. Steinsson. 2014. "Fiscal Stimulus in a Monetary Union: Evidence From US Regions." *American Economic Review* 104, no. 3: 753–792. https://doi.org/10.1257/aer.104.3.753.

Nakamura, E., and J. Steinsson. 2018. "Identification in Macroeconomics." *Journal of Economic Perspectives* 32, no. 3: 59–86. https://doi.org/10.1257/jep.32.3.59.

Pallante, G., E. Russo, and A. Roventini. 2023. "Does Public R&D Funding Crowd-In Private R&D Investment? Evidence From Military R&D Expenditures for US States." *Research Policy* 52, no. 8: 104807. https://doi.org/10.1016/j.respol.2023.104807.

Pesaran, M. H. 2006. "Estimation and Inference in Large Heterogeneous Panels With a Multifactor Error Structure." *Econometrica* 74, no. 4: 967–1012. https://doi.org/10.1111/j.1468-0262.2006.00692.x.

Ramey, V. A. 2011. "Can Government Purchases Stimulate the Economy?" *Journal of Economic Literature* 49, no. 3: 673–685. https://doi.org/10.1257/jel.49.3.673.

Ramey, V. A. 2019. "Ten Years After the Financial Crisis: What Have we Learned From the Renaissance in Fiscal Research?" *Journal of Economic Perspectives* 33, no. 2: 89–114. https://doi.org/10.1257/jep.33.2.89.

Supporting Information

Additional supporting information can be found online in the Supporting Information section. **Data S1**: jae70058-sup-0001-Supinfo.zip.